nature portfolio

Peer Review File

Visualizing Nexus of Porous Architecture and Reactive Transport in Heterogeneous Catalysis by Deep Learning Computer Vision and Transfer Learning

Corresponding Author: Professor Shijie You

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript presents a deep learning method for predicting effective reaction rates in porous catalysts.
The models are trained on synthetic porous media to predict reaction rates calculated with multiphysics simulation, and are applied to optical images of nickel foam electrocatalyst materials yielding good agreement with experimentally measured reaction rate values for a subset of operating conditions.
Finally, there is some discussion of mechanistic interpretation of the effects of certain structural features of the porous media on the resulting reaction rates, which is motivated by visualizations of the activation maps of the regression model.

The research is interesting and seems to be well executed from a technical standpoint, but in my opinion needs to be revised to target a broader general materials science audience given the target journal. The presentation is highly focused on detailed technical questions, and the overall problem setup, approach, and impact is not emphasized strongly enough relative to extensive discussion of detailed comparison of alternative machine learning base models. Furthermore, there seems to be no code or data provided to facilitate reproducibility, and some key details are missing from the methodology (relating to model and training hyperparameters)

* clarity
** overall task framing
One of the main weaknesses of the manuscript is that it lacks clear high-level context for the overall modeling task and goal.

For example, one of the key quantities the paper aims to develop transferable models for is the ratio of the reaction rate to the mass transfer rate Da - but this definition is deferred to the methods section.

It would be much more clean to frame the work as pretraining models on relatively extreme settings of Da (as set up in the beginning of section 2) to enable efficient transfer for any operating conditions.

This motivates the experimental measurement (?) of Da in section 2.3 and the construction of predictive models with a small number of simulations.

> We first determined the Da=0.3 under mass-transfer limitation (c0=1.0 mM L-1) and 1.2 under non-mass transfer limitation (c0=10.0 mM L-1), respectively.

Here, it would be worth discussing the value proposition of the transfer learning workflow vs simply performing the relevant simulations without the machine learning workflow entirely, or compared with just generating a large training dataset at the targeted operating conditions to use to generate the model interpretability maps.

There are several acronyms (e. g. GAS, BET, 3D-CT) that are not defined

** methodology details
*** what position encoding?
The manuscript mentions position tokens, but does not specify what position encoding is used, or how the flow direction is incorporated

*** CNN interpretation
This interpretation of the CNN models

> In brief, the CNN models typically sampled randomly from whole slide images, leading to greater initial uncertainty. Conversely, ViT-FRT began by dividing images into patches based on self-attention mechanisms, which allowed effectively synthesizing information across the full image and more efficient initial learning phase.

*** training efficiency
> These results clearly demonstrated that the ViT-FRT had superior training performance of saving more computing resources

This assertion could be backed by stronger evidence, as the training and evaluation of transformers are commonly much more expensive than CNNs. Conclusions about overall compute resources should factor in the model cost too, not just convergence in terms of training iterations.

*** LSTM variants
> Subsequent to the CNN was a two-layer LSTM for analyzing the sequential data.

> These vectors from the three ResNet branches were concatenated using the torch.cat() function, resulting in a 6144-dimensional feature vector. This comprehensive vector was then introduced to a two-layer LSTM module.

The methods don't seem to explain what is sequential about the data. From the above quoted discussion, the sequence dimension is the channel dimension from the global image representation?
Also, the 3 input channels is a bit confusing and not well motivated. Why not 3D convolutions in this case, for example?

* literature
** There are important gaps in the literature review
*** ML literature
The primary base model used in the paper is Vision Transformer (ViT), but the manuscript does not cite the original ViT publication by Dosovitskiy et al.

Also, the manuscript should also probably should cite the original ResNet report, as well as (probably) the original transformer report by Vaswani et al.

A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale", Proc. Int. Conf. Learn. Representations, 2021.

*** missing some key prior art specifically exploring GANs for generating multiphase and porous structures

*** Bland-Altman analysis needs a citation and a high-level introduction

(Remarks on code availability)
No code or data was provided or linked. as a result, the reproducibility and transparency of the methods is limited.

Reviewer #2

(Remarks to the Author)
This is an elegant piece of work which usefully explores the application of machine learning to heterogeneous catalysis in flow systems using porous materials. The motivation is to predict the overall reaction rate in the system, based on images of the porous material structure. The machine learning systems that have been developed are able to determine the correlation and therefore predict reaction rate for new systems (at specified value of the ratio reaction rate to mass transfer rate, Da) from images of the porous material on 3 axes. The visualisation of the learnt features are enlightening to some extent, and the method applied here (Grad-CAM) shows potential for better physics-informed machine learning tools. The use of the synergy angle also provided some valuable insight into the behaviour of the system.
Although the work has value and I believe it should be published in an appropriate journal, I do not believe it has the level of novelty and significance that would justify its publication in Nature Communications. It is one of a number of works investigating the correlation directly between porous material structure (morphology) and performance properties (such as overall reaction rates). This is a difficult problem and this work makes a useful contribution to it. However, there are some limitations with the work which counter the claims made for it being a paradigm shift. The machine learning systems take 2D images of the structure; to make a step forward for the analysis of complex 3D porous structures, it is necessary to work with the 3D structures, even though this is extremely challenging. The connectivity and tortuosity (amongst other important features) of porous materials are not necessarily captured well in 2D images. The paper states that the use of quantitative structural features (QSFs) for porous materials are not sufficient. But in order to establish the significance of the reported method, it would be necessary to demonstrate that a correlation method between the normalised reaction rate and those

QSFs (and other possibly properties such as Reynolds number and the Da ratio etc.) does not work. This has not been reported. The paper also states that it is using “transfer learning”; but I do not believe the work falls in this category, since it uses the same inputs and outputs (take a structure and predict the reaction rate), and is simply transferred to a different value of one of the variables (Da) (unless I have misunderstood the problem). It would be worth considering using Da as an input to the machine learning model.
It was quite difficult to work out what had been done because the methods are explained in 3 different places (Results, Methods and Supplementary information). I think some information is still missing (such as the inlet flow velocity at each Reynolds number). The presentation of the synergetic angle is poorly explained and needs to consider arbitrary surface orientation and to clearly define all variables.
Some specific points that may be useful in later revisions:
Sec 5, structure generation L423. Growth directions need to be specified quantitatively by angles, in addition to Fig S2 and Tables S1-S2.

L427 and Table S1. Error in definition of cd – the numbers do not match the expression e.g. for P1, 0.0125*epsilon does not equal 0.01125. This error is consistent throughout the table. Should it be (1-epsilon) in each case instead of epsilon?
L492 What is meant by the injection concentration and how does it relate to the initial concentration defined in Table S3?
The boundary conditions need to be defined using axial planes e.g. z=0 for the inflow conditions. The diagram Figure S6 is unclear so this definition is required. The purpose and physical meaning of the symmetry boundary condition and reactive and zero flux boundary conditions should be explained. What is the wider system in which this block fits?
Supplementary text S2 L74-77 refers to some periodic structures and gyroids, referring to Eq 11. This is unclear and does not seem to relate to the rest of the paper.
Supplementary states that the inlet flow velocity is zero, but elsewhere there are 3 different Reynolds numbers.

(Remarks on code availability)
Code was not provided.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I think the revisions adequately address the issues of clarity, broad audience, and significance.

(Remarks on code availability)
The linked repository seems to be complete, and includes both training code and data, as well as a docker file with specific dependency versions.

A README would help provide some context, but the repository is fairly well organized and it straightforward enough to reproduce the paper as is since the main training script is clear.

Reviewer #3

(Remarks to the Author)
The subject of the paper is important, focusing on predicting reaction rates in complex porous media, and challenging, as it considers nontrivial aspects of porous media geometry. In this respect, the paper falls into the category of high-interest and timely subjects that Nature Communications seeks to publish.

The previous reviewers raised several concerns about the paper, with the main ones being (1) that the paper is written in a highly specialized, technical style, and that the overall problem setup, approach, and impact are not emphasized, and (2) that the paper does not possess the level of novelty and significance that would justify its publication in Nature Communications. After reading the revised manuscript, it seems to me that these issues have also persisted in the revised version.

Regarding (1), the authors attempted to improve the presentation in the new version of the manuscript, but I still find it terse and not accessible to a broader audience. For example, in the introduction, the authors emphasize that the primary subject is the exploration of the "tensorial behavior of reaction rate," yet they never clarify what they mean by this. In physical systems, properties can exhibit tensorial behavior if they link two quantities of a vectorial nature. For example, permeability links flow velocity (a vector) to the pressure gradient (also a vector), forming a tensor. Similarly, the dispersion tensor links the concentration gradient (a vector) with particle flux (another vector). In the introduction, when discussing the tensorial character of flow and transport in porous media, the authors reference papers discussing such quantities. However, in what way is the reaction rate a tensor? In all formulations familiar to me, the reaction rate is a scalar. Although it is locally sensitive to boundary conditions, Peclet and Damköhler numbers, etc., due to its derivation as a solution of a partial differential equation, this does not render it a tensorial quantity. From the context, it becomes clear that what the authors mean by the tensorial character of the reaction rate is that it depends on pore geometry in a more subtle manner than through overall porosity; in particular, pore anisotropy does play a role. In my view, these issues are really important for properly understanding the manuscript nad should be expounded upon in the Introduction/Motivation part.

Regarding (2): To characterize the anisotropy, the authors link it to the growth probability of the porous medium in 26 directions. This allows them not only to consider artificially created porous media, but also to reconstruct natural porous media in terms of these growth probabilities. While this approach clearly goes beyond a simple characterization of the porous medium by porosity, I believe it misses several important properties of real porous rocks on which the reaction rate depends. First, the pore space of real rocks has a hierarchical, self-similar character, characterized by a well-defined fractal dimension, as evident from the power-law behavior in scattering profiles (see, e.g., Ref. 5 in the manuscript - Zhang et al., 2020, Figs. 4 and 5 there). Second, the connectivity of the pore space, although downplayed by the authors, is tightly linked with the reaction rate (as only connected pores participate in reactions), and this information does not seem to be encoded in the authors' approach. Third, not only the Damköhler number but also the Peclet number will impact the dissolution patterns, and hence the reaction rate. Finally, particularly at large Damköhler numbers, the dissolution of porous media leads to the formation of wormholes (see, e.g., Kalia and Balakotaiah, Modeling and analysis of wormhole formation in reactive dissolution of carbonate rocks, Chemical Engineering Science 62.4 (2007): 919-928), which completely rebuild both pore space and flow paths in the medium, introducing a new, hierarchical structure of connections. A proper characterization of the reaction rate in such a case is a real challenge in the upscaling of reactive transport. Unfortunately, the authors do not take up this challenge, limiting themselves to relatively homogeneous systems without hierarchical channeling structures. All of these factors limit the applicability of the presented method – it should be perfectly suitable to model total reaction rates in porous catalysts, but the application to dissolving rocks is, in my opinion, rather limited.

Finally, I am also somewhat concerned about the quantity being predicted. R_norm, as defined by Eqs. 12 and 13, is a global quantity, and as such, it is much easier to predict than the local reaction rates at each individual point along the interface. In particular, for large Damköhler numbers, R_norm, by mass conservation, would simply correspond to the total amount of reactant entering the system per unit time (since the aggressive solution does not reach the outlet), and it can be obtained analytically without requiring detailed calculations. At smaller Damköhler numbers, the task is admittedly more challenging, but local reaction rates would seem to be a much more useful quantity to calculate than the overall rate, especially if one is interested in the evolution of the geometry of the reactive system over time.

The limiting factors mentioned above mean that while this paper represents a significant step forward in predicting the properties of reactive porous materials, in my view it does not meet the criteria of innovation and importance required for articles published in Nature Communications. I would suggest publication of these results in a more specialized journal.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
(see the pdf attachment)

(Remarks on code availability)

Version 3:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
I am generally happy with the revisions the authors have made. Perhaps one minor comment: in the new sentence — "However, the current workflow relies on two-phase QSGS datasets, which only captured solid–pore growth at a single scale, but was incapable of reproducing the multiscale, self-organizing evolution of geological porous media, especially in the wormholing regimes [50-52]" — I do not see how references [50] and [52] relate directly to wormholing or multiscale evolution. I suggest keeping there only Ref. 51. Except for this, I believe the paper is ready for publication.

(Remarks on code availability)

I've read with interest the new version of the manuscript, which in my opinion is definitely better than the original version, primarily because the authors were able to recover local reaction rates. This brings the level of novelty and significance expected for *Nature Communications* articles. Most of my previous comments were addressed satisfactorily. The remaining points of disagreement are discussed below (authors' responses from their rebuttal letter are in *italic*):

1. *Added: "The current workflow is not well-suited for more variable porous networks (e.g. non-random pores and geotechnical media), especially those with wormholes and hierarchical features, as the source and transfer datasets generated by the two-phase QSGS method only account for solid/pore growth on a single scale. To develop these applications, the multi-phase QSGS method (e.g. Fast-QSGS61) should be used during dataset construction to capture multiscale pores with self-similar characteristics and complex diagenetic processes."*

Wormholes arise spontaneously due to dissolution-driven self-organization in initially random pore architectures. Thus, the phrase "non-random pores" is overly simplistic and potentially misleading. I suggest making this discussion more nuanced. In particular, it should be stated clearly that the model, in its current form, is not capable of reproducing the evolution of porous media in the wormholing regime.

2. *There are two forms of the Damköhler number: k/U and kL/D. This study used the kL/D expression, which could be rewritten as (UL/D) × (k/U), or Pe × (k/U). Thus, we did not include Pe as a separate input to avoid unnecessary complexity.*

I cannot agree with this reasoning. As is well-established in the literature, dissolution patterns depend on two independent dimensionless numbers: the Peclet and Damköhler numbers - see, for example, Golfier et al., *"On the ability of a Darcy-scale model to capture wormhole formation during the dissolution of a porous medium." Journal of Fluid Mechanics* **457** (2002): 213–254, especially Fig. 12. As shown there, the morphological phase space is two-dimensional and cannot be reduced to a single parameter. Fundamentally, the patterns depend on three physical parameters – the reaction rate (k), diffusion coefficient (D), and fluid velocity (U) – from which two independent dimensionless numbers can be constructed; this is not "*unnecessary complexity*" but a reflection of the underlying physics.. While it is true that Da = Pe × (k/U), expressing it this way eliminates the explicit dependence on U.

The authors should therefore revise this part of the discussion accordingly, but also explicitly address the following question: how is it that their dissolution patterns depend only on the Damköhler number?

Detailed Responses to Reviewers' Comments

Reviewer #1:

Comments: *This manuscript presents a deep learning method for predicting effective reaction rates in porous catalysts. The models are trained on synthetic porous media to predict reaction rates calculated with multiphysics simulation, and are applied to optical images of nickel foam electrocatalyst materials yielding good agreement with experimentally measured reaction rate values for a subset of operating conditions. Finally, there is some discussion of mechanistic interpretation of the effects of certain structural features of the porous media on the resulting reaction rates, which is motivated by visualizations of the activation maps of the regression model.*

The research is interesting and seems to be well executed from a technical standpoint, but in my opinion needs to be revised to target a broader general materials science audience given the target journal. The presentation is highly focused on detailed technical questions, and the overall problem setup, approach, and impact is not emphasized strongly enough relative to extensive discussion of detailed comparison of alternative machine learning base models. Furthermore, there seems to be no code or data provided to facilitate reproducibility, and some key details are missing from the methodology (relating to model and training hyperparameters).

Q1: *The research is interesting and seems to be well executed from a technical standpoint, but in my opinion needs to be revised to target a broader general materials science audience given the target journal.*

Response: The comments are highly appreciated. In response to your suggestion to target a broader general materials science audience, we have extended our focus on predicting and understanding of tensorial behaviors associated with reactive transport for anisotropic porous structures. The importance of this study is summarized as (i) 3DCNN-FRT provided general framework and workflow to visualize the tensorial nexus between arbitrary anisotropic porous architecture and reactive transport. (ii) The trade-off between convenience and accuracy in R_{norm} prediction was realized by reducing operational steps and CPU time compared with numerical simulation and QSFs-based methods. (iii) A common synergy of physical fields was identified to play a pivotal role in diversifying the R_{norm}.

We summarize the innovation of this study as: (i) The isotropy assumption of traditional methods is overcome to enable the capture of tensorial behavior for R_{norm} by multi-modal DLCV framework that integrates directional, structural, and reactive information. (ii) Generality limitation posed by geometric diversity is addressed favored by feature representation transfer learning. (iii) The key structural features are identified for reactive transport, thereby principles of field synergy are proposed by physical visualization of morphological structural features (MSFs). These works are unattainable for conventional QSFs-based methods due to isotropy assumption and abstract nature.

To address this comment, we have re-written the Introduction section that can be

found on Pages 2-3, Lines 41-98. We also added a verification related to mineral dissolution on Page 7, Lines 227-235 in the revised manuscript.

Q2: *The presentation is highly focused on detailed technical questions, and the overall problem setup, approach, and impact is not emphasized strongly enough relative to extensive discussion of detailed comparison of alternative machine learning base models.*

Response: The comments are highly appreciated. We have modified the manuscript to provide more detailed presentation of overall problem setup, approach, and broader impact of the work, whose details can be found in the revised manuscript.

Regarding the overall problem setup, we have made corrections on Page 3, Lines 102-111. Regarding the approach, we have emphasized the strategy and workflow for the problem-solving approach by re-writing the texts and re-drawing Fig. 1 on Page 4, Lines 112-137. Regarding the impact, we have summarized the novelty and impact of this study at the end of revised paper on Pages 13-14, Lines 399-455.

Friendly tips: Since we have modified more than 1200 words totally to address this comment, the revised texts are too long to paste in this document. We kindly invite you to go for details in the places where they are highlighted above in the revised manuscript.

Q3: *there seems to be no code or data provided to facilitate reproducibility, and some key details are missing from the methodology (relating to model and training hyperparameters).*

Response: The code was previously uploaded to the manuscript tracking system, nature communications, but for some reasons that were not apparent, all the reviewers failed to see them when they were reviewing the paper. We are also a little bit confused of what happened to the uploaded codes. Anyway, we are willing to submit these codes again along with the manuscript and relevant documents (*i*. *e*. Appendix codes 1-3). Regarding the methodologies related to model and training hyperparameters, we have made corrections on Pages 16-17, Lines 551-599 in the revised manuscript (709 words modified).

Q4: *overall task framing One of the main weaknesses of the manuscript is that it lacks clear high-level context for the overall modeling task and goal. For example, one of the key quantities the paper aims to develop transferable models for is the ratio of the reaction rate to the mass transfer rate Da - but this definition is deferred to the methods section.*

Response: The comments are highly appreciated. We have revised the manuscript by providing detailed description of the overall modeling task and goal. We have made corrections on Pages 3-4, Lines 102-150 in the revised manuscript (502 words modified).

Q5: *It would be much more clean to frame the work as pretraining models on relatively extreme settings of Da (as set up in the beginning of section 2) to enable*

efficient transfer for any operating conditions. This motivates the experimental measurement (?) of Da in section 2.3 and the construction of predictive models with a small number of simulations.
Response: The comments are highly appreciated. We have further refined the models by pre-training them on a broader range of Da values. Based on the literatures, we have set the Da over the range of 0.01-100 across four orders of magnitude, which are expected to cover the cases from the fastest heterogeneous catalysis to the slowest mineral dissolution. This allows us to develop models that are more representative and more transferable under widespread conditions. This comment also inspires us to further refine the models by optimizing the selection of Da values, images and image order as multi-modal inputs to train the 3DCNN models. This also motivates the experimental measurement of Da in section 2.3 and the construction of predictive models with a small number of simulations. We have made corrections on Pages 3-4, Lines 102-126; and Page 7, Lines 217-235; in the revised manuscript (592 words modified).

Q6: *We first determined the Da=0.3 under mass-transfer limitation (c0=1.0 mM L-1) and 1.2 under non-mass transfer limitation (c0=10.0 mM L-1), respectively.*
Here, it would be worth discussing the value proposition of the transfer learning workflow vs simply performing the relevant simulations without the machine learning workflow entirely, or compared with just generating a large training dataset at the targeted operating conditions to use to generate the model interpretability maps.
Response: The comments are highly appreciated. Followed by Q5, we have added a section to compare the accuracy and convenience of our models with numerical simulation and DLCV models trained on large datasets. The results showed that our models can match the requirement for accuracy of numerical simulation with significant reduction of CPU time. We have made corrections of texts and re-drawn Fig. 4 on Pages 8-9, Lines 245-290 in the revised manuscript (591 words modified).

Q7: *There are several acronyms (e. g. GAS, BET, 3D-CT) that are not defined.*
Response: Agree and corrected accordingly (Page 13, Lines 423-424).

Q8: *what position encoding? The manuscript mentions position tokens, but does not specify what position encoding is used, or how the flow direction is incorporated.*
Response: Followed by Q3, there was something wrong with the codes that were previously uploaded to the manuscript tracking system, nature communications. We are willing to submit these codes again along with the manuscript and relevant documents. Regarding the details in encoding position and flow direction, we have made corrections on Page 3, Lines 107-109; Page 16, Lines 588-591; Page 17, Lines 578-583; in the revised manuscript (220 words modified).

Q9: *The methods don't seem to explain what is sequential about the data. From the above quoted discussion, the sequence dimension is the channel dimension from the global image representation?*

Response: The comments are highly appreciated. We agree that the order of data was indeed represented through the channel dimension, particularly in the 3DCNN models. For our updated 3DCNN models, the six faces of cube-shaped porous structure were concatenated in a specific sequence to form a tensor with the shape (6, 64, 64), where the number 6 represented the channel dimension. Each channel corresponded to one face of the cube, which effectively encoded the image order within the channel dimension. This enabled the 3DCNN models to process images in correct sequence by preserving spatial and positional relationships among the cube's faces during convolutional operation. For Vision Transformer (ViT) models, since input images were concatenated into a three-channel image, the channel dimension did not directly encode the image order. Instead, we used position embeddings to maintain the sequence of the image patches. These position embeddings provided spatial context to each patch, which allowed the models to correctly interpret the order during processing. The complete model descriptions are provided in the response to Q3.

Q10: *Also, the 3 input channels is a bit confusing and not well motivated. Why not 3D convolutions in this case, for example?*
Response: The comments are highly appreciated. Initially, we used three 2D convolutions to represent different faces of the cube. However, this approach was unlikely to fully capture the spatial relationships of structural features. To address this issue, we have modified 2D to 3D convolutions in accordance to your suggestion. The complete model descriptions are provided in the response to Q3. As you can see on Pages 16-17, Lines 551-575, 3D convolution can capture the tensorial behavior of R_{norm} by effectively extracting information on anisotropic and structural growth from 2D images in a tensor-based manner necessary for inferring 3D porous structures. The models focus on high-attention areas such as the inlet boundary, demonstrating its effectiveness in processing 3D structural information.

Q11: *The primary base model used in the paper is Vision Transformer (ViT), but the manuscript does not cite the original ViT publication by Dosovitskiy et al.*
Also, the manuscript should also probably should cite the original ResNet report, as well as (probably) the original transformer report by Vaswani et al.
A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale", Proc. Int. Conf. Learn. Representations, 2021.
Response: Agree. We apologize for our missing citation, and we have added these two important references in the revised manuscript accordingly.

67. A. Dosovitskiy et al. An image is worth 16x16 words: Transformers for image recognition at scale. Proc. Int. Conf. Learn. Representations, (2021)
68. Vaswani, A. Attention is all you need. Advances in Neural Information Processing Systems. (2017).
27. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778 (2016).

Q12: *missing some key prior art specifically exploring GANs for generating multiphase and porous structures*
Response: Agree. We have incorporated important prior works that reported on GANs for generating multiphase and porous structures on Page 13, Lines 429-431 accordingly.

Q13: *Bland-Altman analysis needs a citation and a high-level introduction*
Response: Agree. We have added the appropriate citation and a high-level introduction to Bland-Altman analysis on Page 5, Lines 169-171 in the revised manuscript accordingly.

Reviewer #2:

Comments: *This is an elegant piece of work which usefully explores the application of machine learning to heterogeneous catalysis in flow systems using porous materials. The motivation is to predict the overall reaction rate in the system, based on images of the porous material structure. The machine learning systems that have been developed are able to determine the correlation and therefore predict reaction rate for new systems (at specified value of the ratio reaction rate to mass transfer rate, Da) from images of the porous material on 3 axes. The visualisation of the learnt features are enlightening to some extent, and the method applied here (Grad-CAM) shows potential for better physics-informed machine learning tools. The use of the synergy angle also provided some valuable insight into the behaviour of the system.*

Q1: *I do not believe it has the level of novelty and significance that would justify its publication in Nature Communications. It is one of a number of works investigating the correlation directly between porous material structure (morphology) and performance properties (such as overall reaction rates).*
Response: The comments are highly appreciated. To enhance the novelty and significance, we have made throughout modifications from Introduction, Methodologies, to Results and discussion. We have modified our focus from predicting reaction rate in heterogeneous catalysis to predicting and understanding of tensorial behaviors associated with reactive transport for anisotropic porous structures.

The novelty of this study is summarized as: (i) The isotropy assumption of traditional methods is overcome to enable the capture of tensorial behavior for R_{norm} by multi-modal DLCV framework that integrates directional, structural, and reactive information. (ii) Generality limited by geometric diversity is addressed favored by feature representation transfer learning. (iii) The key structural features are identified for reactive transport, thereby principles of field synergy are proposed by physical visualization of morphological structural features (MSFs). These works are unattainable for conventional QSFs-based methods due to isotropy assumption and abstract nature.

The importance of this study is summarized as: (i) 3DCNN-FRT provided general framework and workflow to visualize the tensorial nexus between arbitrary anisotropic porous architecture and reactive transport. (ii) The trade-off between convenience and accuracy in R_{norm} prediction was realized by reducing operational steps and CPU time compared with numerical simulation and QSFs-based methods. (iii) A common synergy of physical fields was identified to play a pivotal role in diversifying the R_{norm}.

Friendly tips: Since we have modified more than 4200 words totally to address this comment, the revised texts are too long to paste in this document. We kindly invite you to go for details in the places where they are highlighted below in the revised manuscript.

Q2: *The machine learning systems take 2D images of the structure; to make a step forward for the analysis of complex 3D porous structures, it is necessary to work with the 3D structures, even though this is extremely challenging. The connectivity and tortuosity (amongst other important features) of porous materials are not necessarily captured well in 2D images.*

Response: The comments are highly appreciated. Compared with 2D images, 3D structures are more favorable for calculating R_{norm} in porous media, but challenges remain due to high cost in sampling, complexity in identifying key structural features and inefficiency in handling large datasets.

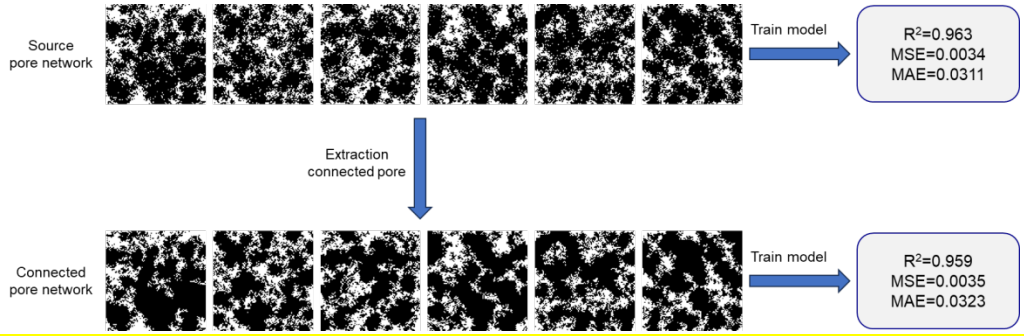
In response to your comment, we have incorporated a 3D Convolutional Neural Network (3DCNN), as it is well-suited for handling 3D structures. Owing to dependent on 2D images of cube's faces, the 3DCNN processes these images in a tensor-based manner by capturing the structural growth and anisotropy in 3D porous media. This allows inferring 3D structures from 2D images [62] by overcoming the limitation of traditional methods that assume isotropy and rely on abstract statistical metrics.

In our revised manuscript, the updated models integrate structural (MSFs across six faces), directional (image order), and reactive (Da) information to predict the tensorial behavior of R_{norm}. The results showed that the new models could capture both tensorial and random variation in R_{norm}, which traditional QSFs methods failed to achieve. Furthermore, the models focus on high-attention areas such as the inlet boundary, demonstrating its effectiveness in processing 3D structural information.

To assess the impact of connectivity, we have trained additional models upon the source and transfer datasets after filtering out non-connected pores using the "bwconncomp" function in MATLAB (Supplementary Fig. 17, details in the **Appendix code 2**). Changes in connectivity and tortuosity had negligible impact to the R^2, MSE, or MAE during training (details in the **Appendix code 3**), indicating that the 3D CNN relied on fundamental structural anisotropy rather than QSF metrics. The 3D CNN performed well in the absence of complete connectivity and tortuosity.

Taken together, while the input images may lack some important QSFs like tortuosity and connectivity, the 3DCNN can effectively capture the anisotropic and structural growth information necessary for inferring 3D porous structures. This

makes it feasible for predicting R_{norm} without relying on QSFs based on isotropic assumption.



Supplementary Fig. 17 | Model performance before and after extracting connected pores.

We have made corrections on Pages 3-4, Lines 102-137; Pages 8-9, Lines 245-290; Pages 16-17, Lines 551-575 in the revised manuscript (1300 words modified).

Q3: *But in order to establish the significance of the reported method, it would be necessary to demonstrate that a correlation method between the normalised reaction rate and those QSFs (and other possibly properties such as Reynolds number and the Da ratio etc.) does not work.*

Response: The comments are highly appreciated. In our revised manuscript, we have provided three scenarios where MSFs work but QSFs fail to adequately capture the behavior of R_{norm}, they are: (i) capturing the tensorial behavior of R_{norm} in single anisotropic porous structure, (ii) capturing the random variation of R_{norm} across multiple anisotropic porous structures that share similar QSFs but differ in their anisotropies, and (iii) identifying geometric structures responsible for R_{norm}.

To address issues, we have added relevant methods, results, discussion and figures on Pages 8-9, Lines 245-290, Pages 12-14, Lines 399-455 and Pages 16-17, Lines 551-575 in the revised manuscript. (1635 words modified). Added contents are enclosed below.

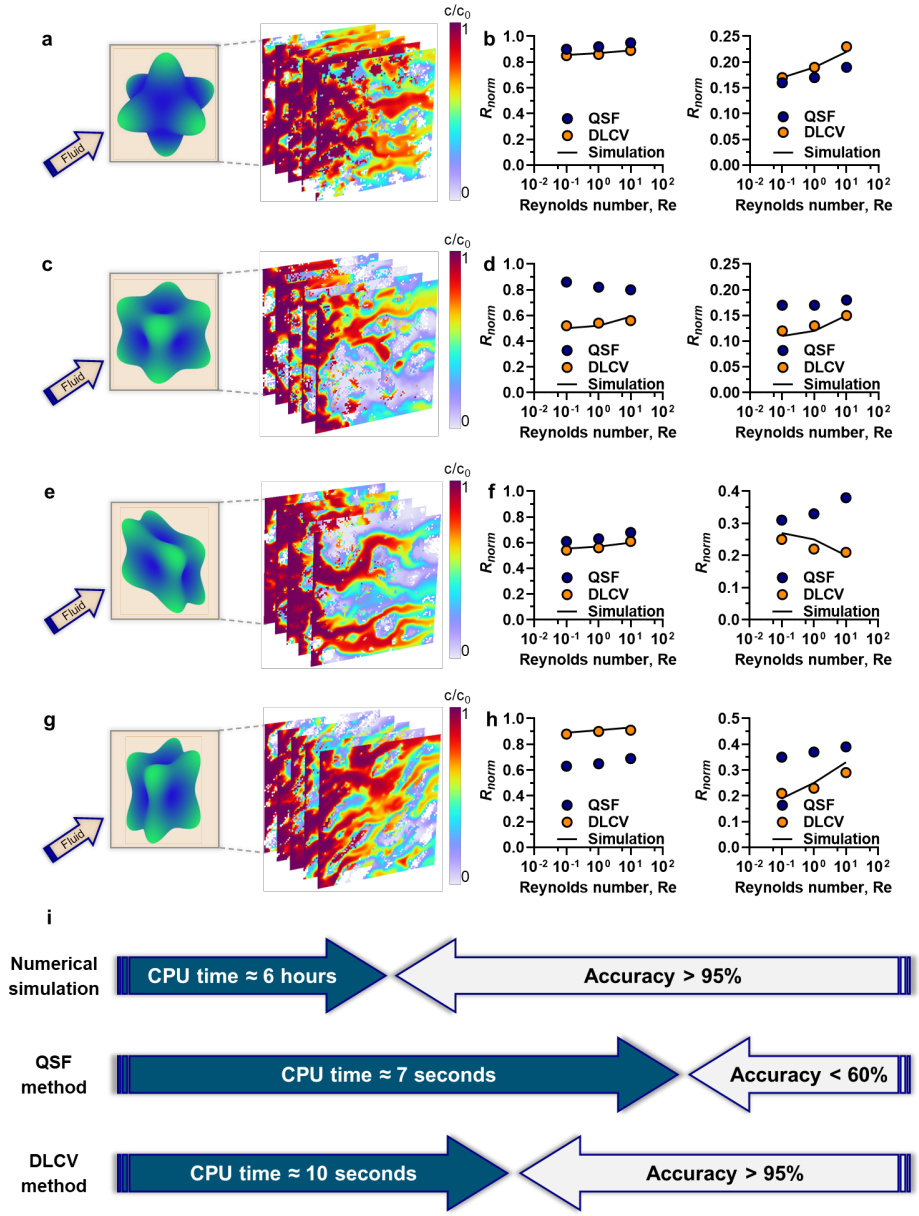


Fig. 4 | Accuracy and convenience comparison for different methods. Concentration distribution for **a,** Axis anisotropic and **b,** Diagonal anisotropy architectures. Accuracy comparison for **c,** Axis anisotropic and **d,** Diagonal anisotropy architectures with similar QSFs. **e,** Concentration distribution and **f,** Tensorial R_{norm} variations for axis anisotropic architecture. **g,** Concentration distribution and **h,** Tensorial R_{norm} variations for diagonal anisotropic architecture. and **i,** Relationship between CPU time and accuracy for different methods.

Pages 8, Lines 245-249: 2.4 Comparison of 3DCNN-FRT, QSFs and numerical simulation methods

We assessed the predictive accuracy and convenience of 3DCNN-FRT, QSFs method and numerical simulation by examining tensorial behavior of both intra- and inter-structures. The ML model reported by Liu et al.[18] was used to represent QSFs-method due to high prediction accuracy and extensive training data.

Page 8, Lines 250-271: Accuracy: We proposed two kinds of hypothetical structures (named as HA and HB) being similar as QSFs but with different g_i values,

in which HA and HB had higher g_i values along diagonal and principal axes directions, respectively (Supplementary Fig. 8). Numerical simulations revealed a significant difference in R_{norm} between HA and HB. For instance, compared to HA whose R_{norm} was 0.57 and 0.42 for 0.3 and 0.7 porosity at Re=0.1, R_{norm} for HB was reduced by half (0.33 and 0.24 for 0.3 and 0.7 porosity; Figs. 4a-d). Notably, the QSFs-method yielded similar results for HA (R_{norm} = 0.52 and 0.38) and HB (R_{norm} = 0.55 and 0.39), indicating its inability to capture R_{norm} diversity across structures with similar QSFs but different structural anisotropy. In contrast, 3DCNN could predict the R_{norm} with an accuracy up to 95.6%.

We also assessed the accuracy in capturing the inter-structured tensorial behavior of R_{norm} for anisotropic structure (HC) with diagonal textures on x-z plane (Supplementary Fig. 8). We simulated the R_{norm} values under x+ fluid and y+ fluid conditions defined in Fig. 1 at 10 Da. As shown in Figs. 4e-4h, R_{norm} varied significantly with fluid direction. When the fluid direction was changed from x+ to y+, R_{norm} shifted from 0.72 and 0.65, to 0.56 and 0.41, for 0.3 and 0.7 porosity at Re=0.1. The QSFs-method failed to capture the tensorial behavior of R_{norm} (Fig. 4f) because most critical factors (*e. g.* specific surface area, pore orientation and even coordination number) were scalars rather than vectors, in which only tortuosity and pore-throat ratio were adjusted according to fluid direction[54]. Notably, 3DCNN-FRT could achieve an accuracy of 97.6% (Fig. 4h). This should be due to the unique ability of multilayer neural networks to directly learn information on structural growth in 3D space of images[30, 42], rather than relying on statistical metrics like connectivity (quantified by coordination number) and tortuosity. This suggested the robustness of 3DCNN-FRT in predicting tensorial behaviors of R_{norm} for reactive transport within anisotropic porous structures.

Pages 8-9, Lines 272-290: Convenience: 3DCNN-FRT was more convenient than both numerical simulation and QSFs method (Fig. 4i). First, unlike hard-to-obtain 3D physical models and QSFs information, it was much easier for 3DCNN-FRT to acquire binarized images by using standard camera and image processing software (*e. g.* ImageJ)[55]. Second, 3DCNN-FRT could generate valuable results within short time of 10 seconds by simple input command, whereas numerical simulation required longer than 6 hours to produce results for single case. Third, 3DCNN-FRT workflow enabled rapid establishment of structure-property relationships for new anisotropic structures by only 40 iterations and 50 data pairs (Fig. 4i). This resulted in approximately 2000-fold reduction in CPU time compared with traditional methods. In summary, 3DCNN-FRT not only matched the prediction accuracy of numerical simulation, but also boosted convenience by reducing time, cost and operational complexity in practical applications.

Q4: *The paper also states that it is using "transfer learning"; but I do not believe the work falls in this category, since it uses the same inputs and outputs (take a structure and predict the reaction rate), and is simply transferred to a different value of one of the variables (Da) (unless I have misunderstood the problem). It would be worth considering using Da as an input to the machine learning model.*

Response: The comments are highly appreciated. We have made throughout modifications to the contents relevant to transfer learning. In the revised manuscript, Da appears as an input variable to form a multi-modal input alongside MSFs and image order. This allows the model to leverage Da directly in its predictions. Being different from prior version, transfer learning is now used to address the generality limit posed by the diversity of anisotropic porous structures. We believe this to be a well-recognized usage for transfer learning. For example, Sluilt et al.[51] used transfer learning strategy to predict droplet growth rates in microfluidic devices with various channel structures (*Nature Communications*, 2021, 12(1): 25.). Herein, we for the first time apply this approach to address generality challenges caused by diversity of porous structures.

In brief, we categorize reactive transport into extreme isotropic cases (*i. e.* source task) and arbitrary anisotropic cases (*i. e.* transfer task). Transfer learning allows extracting fundamental structure-property relationship for isotropic case without interference of structural anisotropy, which enables precise fine-tuning of transfer model for anisotropic case.

We have made corrections on Pages 3-4, Lines 102-137, and Pages 6-7, Lines 195-243 in the revised manuscript (1109 words modified). Added contents are enclosed below.

Pages 3-4, Lines 102-137: To capture the tensorial behavior of reaction rate, identify the key reaction sites for reactive transport, and address the generality limitation posed by geometric diversity, we established the transferable 3DCNN-TL framework to map structural, directional and reactive information to normalized effective reaction rate (R_{norm}) that reflected the relative magnitude between effective reaction rate and well-mixed rate[40, 41]. The structural information consisted of images (256×256 pixels) from six faces of the cube-shaped porous architecture. The image order encoded directional information corresponding to upstream, top, bottom, two sides, and downstream, as neural networks could interpret the sequence of input images (Supplementary Fig. 1)[42]. And the reactive information was described by dimensionless Damköhler number (Da) that represented the ratio of solid-liquid reaction rate to solute diffusion rate[43].

To enable rationality of TL strategy in terms of completeness of dataset, robustness and operability of workflow, the reactive transport was categorized into isotropic cases with specific reaction modes (*i. e.* source task) and anisotropic cases with arbitrary modes (*i. e.* transfer task; Figs. 1a-c). This framework allowed extracting fundamental structure-property relationship without interference of structural anisotropy, followed by fine-tuning of transfer models for anisotropic cases based on small dataset. To achieve this goal, we regulated the structural anisotropy using quartet structure generation set (QSGS)[44] method to construct porous structures with desired porosity ε, core distribution probability c_d, and growth probability in 26 directions g_i (i =1,2,...,26; Figs. 1a and c, Method and Supplementary Text 1 and Fig. S2). For isotropic cases, we set the uniform g_i (0.001) in 26 directions and selected 5 values of Da=0.01, 0.1, 1, 10, and 100 [23, 45-47]. For anisotropic cases, the g_i values were arbitrarily combined, and Da values were randomly selected in the range of

0.01-100. The source dataset contained 18000 pairs of data, and the transfer dataset had only 50-140 pairs of data (Detail in Method). According to analysis for R_{norm} data structure shown in Supplementary Text 2 and Fig. S3, we fine-tuned the source models by using feature representation transfer learning (FRT)[48]. Theoretically, the TL strategy could potentially cover all the reactive transport scenarios.

We established a 3DCNN-FRT workflow for custom users, including base process and custom process. For base process (Fig. 1d), the source models were trained on source database to learn fundamental structural-property relationships for isotropic porous structures under extreme Da values. For the custom process (Fig. 1e), the g_i values in 26 directions were first calculated from the images of porous structure using the method described by Chiang et al.[49], such that the transfer dataset could be generated. The source model was fine-tuned upon the small-scale dataset to allow buildup of custom structure-property relationships for target porous structures. Last, the salient features for both isotropic and anisotropic porous media were identified, and the hydrodynamic mechanisms could be derived from synergy of physical fields. This framework would be applied for predicting and elucidating reactive transport in various anisotropic porous structures such as compacted soil[50], porous catalysts[51] as well as fractured rocks[52].

Pages 6-7, Lines 195-243: The DLCV models are sensitive to heterogeneity of the training and test data. The model trained with images-result pairs collected under one condition is unlikely to be extended for other conditions. To evaluate the transferability of 3DCNN-FRT, we next evaluated the transfer training performance and predicting precision for 3DCNN-FRT transfer model that was constructed by FRT strategy (Fig. 3a). Specifically, all other layers were frozen except for that being fully connected layer to prevent the transfer model from overwriting previous features and parameters[38]. Then, the entire model was fine-tuned to boost the performances. This model was trained over 150 epochs based on anisotropic transfer dataset with 60 images-result pairs characterized by nonuniform g_i and 0.1 Da (Fig. 3a and Supplementary Fig. 5), which was not seen by the source model before.

Notably, the 3DCNN-FRT transfer model held good training performance in terms of 73% reduction in loss over 20 epochs. We then tested predicting precision of the transfer model using 1000 untrained data, in which both pore structures and sizes were totally different from that in transfer databases. The 3DCNN-FRT predicted the R_{norm} values with an accuracy as high as 98.2%, whereas training the 3DCNN from scratch on the small-scale transfer dataset resulted in overfitted model with accuracy as low as 23.6% (Fig. 3b and Supplementary Fig. 6). Mean difference and standard deviation for 3DCNN-FRT transfer remained 0.031 ± 0.0372 (Fig. 3c), suggesting excellent transferability of 3DCNN-FRT in predicting R_{norm} across anisotropic porous architectures. We also plotted the salient features identified by transfer model in Fig. 3d, where specific pore structures were also highlighted with clearly defined boundaries. This opened up opportunities to further explore the common mechanisms driving R_{norm} variations in both isotropic and anisotropic porous architectures.

We next evaluated whether 3DCNN-FRT could be applied for R_{norm} prediction in

electrochemical conversion of $K_3[Fe(CN)_6]$ on foam nickel (Figs. 3e-g) using the knowledge learned from pre-trained databases. The selection of nickel electrode resulted from its high catalytic activity and selectivity for $Fe(CN)_6^{3-}$ reduction, which was extensively verified in previous studies[45]. We first set Da=0.3 under non-mass-transfer limitation (c_0=1.0 mM L^{-1}) and Da=1.2 under mass transfer limitation (c_0=10.0 mM L^{-1}), respectively. Then, the transfer learning models could be built to predict R_{norm} values with six representative square images of foam-structured nickel (Fig. 3f). As shown in Fig. 3g, 3DCNN-FRT showcased a great performance in predicting the tested R_{norm} values for electrochemical conversion of $K_3[Fe(CN)_6]$ under different Re numbers and fluid directions with error not exceeding 7.1%.

3DCNN-FRT was further tested on natural carbonate dissolution reported by Menke et al.[21], which posed a threat to long-term security of carbon capture and storage (CCS). Dissolved CO_2 formed HCO_3^- solution that reacted with the carbonate, resulting in carbonate dissolution and CO_2 escape. This reaction occurring at solid-liquid interface followed the equation shown in Fig. 3h, at Da=0.06. The carbonate had a larger g_i in the horizontal direction than that in the vertical direction (Fig. 3i and Supplementary Fig. 7). 3DCNN-FRT predicted R_{norm} values with 96.4% accuracy in accordance to custom workflow. These results suggested the feasibility of obtaining high-performance structure-R_{norm} relationships by 3DCNN-FRT source model and transfer model, without the need to develop new models from scratch.

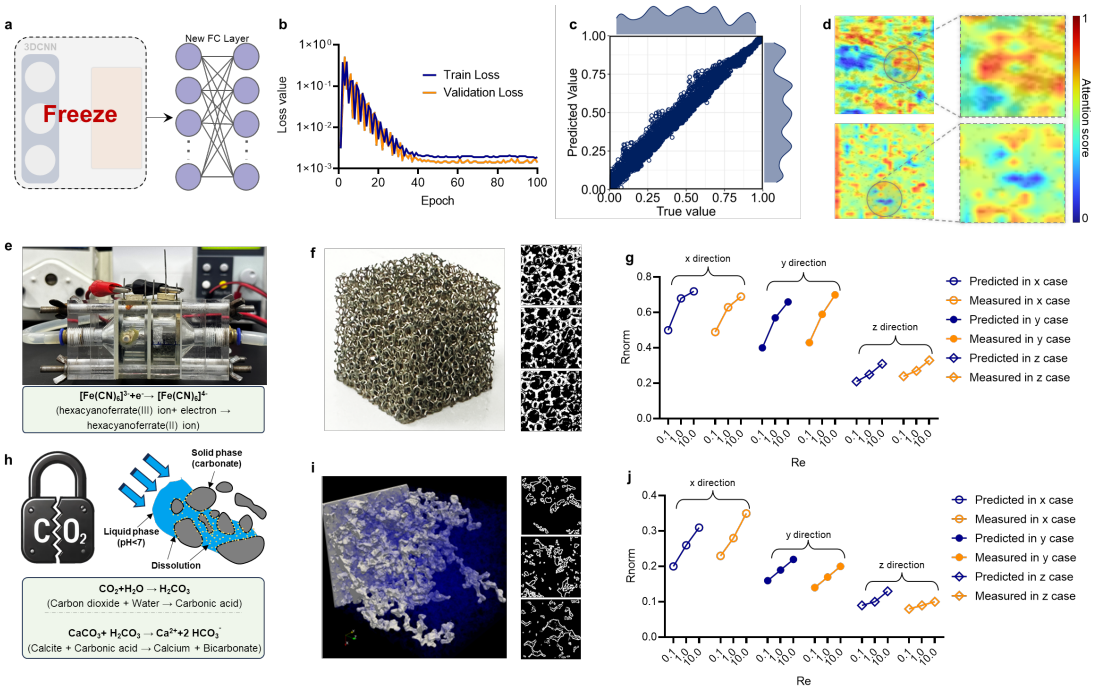


Fig. 3 | Transfer Learning Performance. a, Schematic representation of FRT strategy for 3DCNN. **b,** Learning curves. **c,** Predicted results. **d,** heat map for FRT strategy. **e,** Experimental setup for electrochemical conversion of $K_3[Fe(CN)_6]$ on foam nickel. **f,** Nickel foam electrode and representative images. **g,** Comparative analysis between tested and predicted R_{norm} under different Re and fluid directions. **h,** Schematic for carbonate dissolution. **i,** carbonate structure and representative images. **j,** Comparative analysis between tested and predicted R_{norm} for carbonate dissolution.

Q5: *It was quite difficult to work out what had been done because the methods are explained in 3 different places (Results, Methods and Supplementary information). I think some information is still missing (such as the inlet flow velocity at each Reynolds number).*

Response: The comments are highly appreciated. We have added detailed descriptions of methods on Pages 14-17, Lines 457-599 in the revised manuscript. Added contents are enclosed below.

Page 14, Lines 457-489: Input Data Generation.

To simplify the setup, we focused on cube-shaped porous architectures with unidirectional flow perpendicular to the surface. The six faces of the porous architecture were defined as upstream, top, bottom, two sides, and downstream based on the flow direction (Supplementary Fig. 1). The cube porous structures were generated by quartet structure generation set method[61], in which there were three major input parameters, *i. e.* porosity ε, core distribution probability c_d, and growth probability in 26 directions g_i (i =1,2,...,26). The c_d was used to control the averaged pore size, and the g_i was used to control the anisotropy of the pore shape. Supplementary Fig. 2 and Table 1 showed the 26 growth directions of the pores. Supplementary Fig. 1 provided an example of the generated porous architecture model, where the pores (white phase) were stochastically distributed within the continuous solid phase (black phase).

For source dataset with isotropic porous structures, the porosities ε were set to be 10%, 30%, 50%, 70% and 90%. The pore sizes were controlled by taking different core distribution probability c_d, including $0.0125(1-\varepsilon)$, $0.025(1-\varepsilon)$, $0.05(1-\varepsilon)$ and $0.1(1-\varepsilon)$. Structural isotropy was achieved by setting growth probability in 26 directions g_i to 0.001. Each parametric combination was explored, generating 10 iterations per cluster. Considering each structure had 6 inlet directions, 5 kinds of Da values (*i. e.* 0.01, 0.1, 1, 10, 100) and 3 kinds of Re values (*i. e.* 0.1, 1, 10), this resulted in the synthesis of 18000 scenarios (5×4×10×6×5×3).

For transfer dataset with anisotropic structures, the porosities ε and core distribution probability c_d were consistent with those in the source dataset. The growth probability in 26 directions g_i was adjusted according to the images of target structures based on Slice-DAC method reported by Kench et al.[52]. Each parametric combination was generated only once. The $0.0125(1-\varepsilon)$ and $0.1(1-\varepsilon)$ cases were designated for fluid flow in the x+ direction, while $0.025(1-\varepsilon)$ and $0.05(1-\varepsilon)$ cases were assigned to the y+ and z+ directions, respectively. Considering each structure had 1 inlet directions, 1 to 5 kinds of Da values and 3 kinds of Re values, this resulted in the synthesis of 60 to 300 scenarios (5×4×1×1×[1 to 5]×3).

Notably, the range of characteristics length within these porous media extended from 10^{-5} to 10^{-3}, serving as a limitation on the available image information. Subsequently, six lateral images were extracted from each porous geometric model to serve as input structural information (Supplementary Fig. 1). The order for these images was used to encode directional information, corresponding to the upstream,

top, bottom, two sides, and downstream. And Da value was used as reactive information.

Pages 16-17, Lines 551-599: Model architecture

To select the most robust model for 3D structure processing, we compared two models with different logics. The first was 3DCNN, which directly processed 3D structures by stacking multiple images into a tensor in 3D space. The second model was ViT, which indirectly processed 3D structures by dividing 2D images into patches and analyzing their positional relationships in 2D space. Since 3DCNN and ViT could interpret the order of input images, we used image order to encode flow direction. Consequently, both models were trained with the same inputs: six grayscale images from the cube surfaces, the image order, and the Da value.

3DCNN: For 3DCNN model, the six images were stacked in the order of upstream, top, bottom, two sides, and downstream into a 3D tensor with dimensions (6, 64, 64), where 6 was the channel number. These six channels represented the six faces of the cube, and their order was critical for the model's learning and predicting. Then, the 3D tensor was fed into the model, which consisted of three 3D convolutional layers, three 3D max-pooling layers, and two fully connected layers. The first convolutional layer took in the six input channels and outputted 16 channels, followed by 3D max-pooling. The second and third convolutional layers outputted 32 and 64 channels, respectively, with each followed by pooling. After these layers, the tensor was flattened and mapped into a 64-dimensional space via a fully connected layer. Additionally, an embedding layer was used to map the Da value into the same 64-dimensional space, which was then combined with the image sequence information to better capture the features. Finally, the convolutional features, image sequence information, and Da embedding were passed through a final fully connected layer to produce the three target outputs (Supplementary Fig. 15).

The key hyperparameters of the 3D CNN model included as follows. The convolutional layers had a kernel size of 3×3×3 with 16, 32, and 64 channels; the pooling layers used a 3×3×3 kernel with a stride of 2; and the Da value was embedded into a 64-dimensional space. The learning rate was set to 0.001, and the model was trained for 150 epochs.

Vision Transformer (ViT): For ViT model, the six grayscale images were combined in the same order as in 3DCNN to form a three-channel image with dimensions of (224, 224, 3), where 3 was the channel number. The model's Patch Embedding module then divided this image into 32×32 patches, which were flattened into a sequence. A positional encoding was added to each patch's embedding vector, allowing the model to recognize the relative positions of patches within the original image. The core of the ViT model consisted of multiple encoder blocks, each comprising a Multi-Head Self-Attention mechanism and a Feed-Forward Network (FFN). Within each encoder block, the embedded patches were first normalized, then processed by the Multi-Head Self-Attention mechanism, followed by feature extraction through the FFN. The Da value was mapped to a 64-dimensional vector via a linear layer and then concatenated with the image features extracted by ViT (CLS

Token). This combined feature vector was passed through a fully connected layer to produce the three target outputs (Supplementary Fig. 16).

The key hyperparameters of the ViT model included as follows. A patch size of 32×32, an embedding dimension of 768, 3 encoder blocks, 8 attention heads, and an FFN expansion ratio of 4. The Da value was embedded into a 64-dimensional space. The learning rate was set to 0.001, and the model was trained for 150 epochs.

Page 14, Lines 533-534: Inlet flow velocity for each Re value was calculated by $Re\mu/L$, L was characteristic length calculated by π/s_A, and s_A was the specific surface area.

Q5: *The presentation of the synergetic angle is poorly explained and needs to consider arbitrary surface orientation and to clearly define all variables.*

Response: The comments are highly appreciated. The principle of physical field synergy (PFS) was first reported by Guo et al. (*Int. J. Heat & Mass Tran.*, 2005, 48(9): 1797-1807) in the field of heat transfer. Based on Fourier's Law and equations of heat flux within boundary layer, they found that convective heat transfer could be enhanced by reducing the angle between flow field and temperature gradient. This indeed offered an extremely simple manner for enhancing heat transfer. Inspired by this finding and basic knowledge on analogy between momentum, heat and mass transfer, we propose PFS principle orienting mass transfer based on Fick's Law and equations of heat flux within boundary layer (Yu and You, et al., *Adv. Funct. Mater.*, 2023, 33(26): 2214725), which can offer useful tools for explaining and enhancing mass transfer process. Since mass transfer is the key process involved in reactive mass transport, we assume that PFS may have a role in this study. We have added detailed descriptions of PFS in elucidating the nexus between anisotropic porous architecture and reactive transport by deep learning computer vision and transfer learning on Pages 11-12, Lines 329-372 in the revised manuscript. Added contents are enclosed below, hoping that the resulting clarity has improved the manuscript.

Pages 10-11, Lines 322-365: To verify above hypothesis, we attempted to elucidate the physical fields within the salient pore throats and curved channels. According to Eq. 12, R_{norm} was linked to two variables, *i. e.* reaction rate constant for fluid-solid interfacial reaction (k_r) and local concentration (c_i). Since k_r was constant, c_i was the dominant variable that relied on mass transfer. To identify the dominant factors in local mass transfer, we integrated the governing equations (Eq. 8 in Materials and Methods) over the solid-liquid interface as

(1)

Nondimensionalization produced the correlation as

(2)

and further we had

(3)

In arbitrary geometries, the local synergic angle was written as

(4)

where Sh, Re, Sc was the local Sherwood number (Sh=$k_m x/D$), local Reynolds number (Re=$\rho u x/\mu$) and Schmidt number (Sc=$\mu/\rho D$). u was the local velocity vector at any point on the surface. ∇C was the concentration gradient vector that varied spatially within the porous structure. θ was the local synergic angle that measured the alignment of two vectors. Calculation of θ relied on dot product between ∇C and u vectors, in which θ was the scalar being **independent on surface orientation,** geometric configuration and external reference frame. It was only the function of relative direction of the two vectors at any given point, making it applicable to the surface with arbitrary orientations.

Eq. 1 suggested three pathways available for enhancing mass transfer, and they were (i) increasing Re value by increasing flow velocity, (ii) increasing μ by using high-viscosity fluid, and (iii) increasing dimensionless integration (*i. e.* St, Sum of the mass sources within boundary layer) by increasing the value of cosθ, concentration gradient and velocity. Among them, minimizing θ allowed the maximization of cosθ and Sh_x value (Fig. 5d), giving rise to what is termed principle of physical field synergy (PFS, θ is the synergic angle, St is synergic number)[57, 58]. The PFS was adopted in the field of heat and mass transfer enhancement.[59, 60] Evident is that smaller θ led to higher rate of reactant transport onto solid surface by water flow (Eqs. 3-4). To verify the effect of PFS on R_{norm}, we performed the correlation analysis between R_{norm} and average synergic angle θ_V calculated by volumetric averaging of local synergic angles as

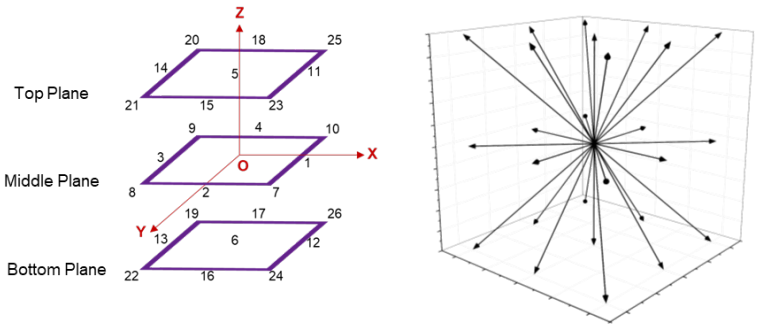
(5)

where V was the total porous volume, V_j and θ_j the volume and synergic angle of the j_{th} pore unit, and n the total number of units at flow section.

Q6: *Sec 5, structure generation L423. Growth directions need to be specified quantitatively by angles, in addition to Fig S2 and Tables S1-S2*

Response: The comments are highly appreciated. We have modified the growth direction by providing a clear explanation of each direction including the quantitative specification of growth directions by azimuthal and elevation angles. We have added detailed descriptions of methods on Page 4, Lines 142-144 in the revised manuscript. Added contents are enclosed below.

Page 4, Lines 142-144: Structural anisotropy was obtained by changing g_i in 26 directions (Figs. 1a and b), which corresponded to the vectors from the cube center to each edge midpoint, vertex, and face center. The direction 1, 2 and 5 represented x, y and z direction in coordinate system, respectively.



Supplementary Fig. 2 | Diagram of the 26 growth directions: from the center of the cube to each edge midpoint, vertex, and face center. x+ (direction 1), x- (direction 3), y+ (direction 2), y- (direction 4), z+ (direction 5), and z- (direction 6). Azimuthal angle and elevation angle for each direction were given in Table S1.

Table S1 Azimuthal angle and elevation angle for 26 directions in Supplementary Fig. 2.

Direction Number	Direction (x, y, z)	Aziumth (°)	Elevation (°)
1	(1, 0, 0)	0	90
2	(0, 1, 0)	90	90
3	(-1, 0, 0)	180	90
4	(0,-1, 0)	-90	90
5	(0, 0, 1)	0	0
6	(0, 0, -1)	0	180
7	(1, 1, 0)	45	90
8	(-1, 1, 0)	135	90
9	(-1, -1, 0)	-135	90

10	(1, -1, 0)	-45	90
11	(1, 0, 1)	0	45
12	(1, 0, -1)	0	135
13	(-1, 0, -1)	180	135
14	(-1, 0, 1)	180	45
15	(0, 1, 1)	90	45
16	(0, 1, -1)	90	135
17	(0, -1, -1)	-90	135
18	(0, -1, 1)	-90	45
19	(-1, -1, -1)	-135	125.3
20	(-1, -1, 1)	-135	54.8
21	(-1, 1, 1)	135	54.7
22	(-1, 1, -1)	135	125.3
23	(1, 1, 1)	45	54.7
24	(1, 1, -1)	45	125.3
25	(1, -1, 1)	-45	54.7
26	(1, -1, -1)	-45	125.3

*Azimuthal angle (θ_{AA}) is the angle between the projection of the vector onto the xy-plane and the positive x-axis, defined as θ_{AA}=arctan(y/x);
*Elevation angle (θ_{EA}) is the angle between the vector and the positive z-axis, defined as θ_{AA}=arccos(z/r). Where r is the magnitude of the vector.

Q7: *L427 and Table S1. Error in definition of cd – the numbers do not match the expression e.g. for P1, 0.0125*epsilon does not equal 0.01125. This error is consistent throughout the table. Should it be (1-epsilon) in each case instead of epsilon?*
Response: Agree and corrected accordingly. The corrected expression is (1-ε). We have made the necessary corrections throughout the revised manuscript. Corrected contents are enclosed below.

Page 14, Lines 470-471: The pore sizes were controlled by taking different core distribution probability c_d, including 0.0125(1-ε), 0.025(1-ε), 0.05(1-ε) and 0.1(1-ε).

Q8: *L492 What is meant by the injection concentration and how does it relate to the initial concentration defined in Table S3?*

Response: The term "injection concentration" is referred to as the "initial concentration" in Table S3. To clarify this point, we have revised the manuscript to standardize the terminology by replacing "injection concentration" with "initial concentration" to enable consistence with Table S3. Corrected contents are enclosed below.

Page 14, Lines 546-549:

(13)

where c_0 was the initial concentration. R_{norm} was the normalized effective reaction rates that quantified the discrepancy between R_r and the well-mixed reaction rate for $c=c_0$.

Q9: *The boundary conditions need to be defined using axial planes e.g. z=0 for the inflow conditions. The diagram Figure S6 is unclear so this definition is required. The purpose and physical meaning of the symmetry boundary condition and reactive and zero flux boundary conditions should be explained. What is the wider system in which this block fits? Supplementary text S2 L74-77 refers to some periodic structures and gyroids, referring to Eq 11. This is unclear and does not seem to relate to the rest of the paper.*

Response: The comments are highly appreciated. We have rewritten the boundary conditions to clearly define them by using a coordinate system as

(1) Boundary conditions:

Based on the comment raised by Reviewer #1, we have extended our focus from predicting reaction rate in heterogeneous catalysis to predicting and understanding of tensorial behaviors associated with reactive transport for anisotropic porous structures. To this end, we have modified the boundary conditions by replacing the original symmetric boundaries with periodic boundaries. This means that the fluid or substance exits one boundary and re-enters the opposite boundary in the same state, so as to effectively simulate continuous flow in the porous structure and represent the infinite expansibility of media. We have made corrections on Pages S4-S6, Lines 85-132 in the revised manuscript. Added contents are enclosed below.

Pages S4-S6, Lines 85-132: The initial values of pressure and flow velocity were set as 0 Pa and 0 m s^{-1}, respectively. The initial concentration was c0. Moreover, the boundary conditions can be classified as inflow boundary, outflow boundary, periodic boundaries, and reactive and zero-flux boundaries. To clarify the boundary conditions and their directions, we first defined the coordinate axes based on the anisotropic growth directions:

The origin is set at the midpoint of the cube, with the positive x-axis aligned with growth direction 1, the positive y-axis with direction 2, and the positive z-axis with

direction 3. The axis dimensions correspond to the number of pixels. This setup results in six possible flow directions perpendicular to the cube's faces: x+ (along direction 1), x- (along direction 3), y+ (along direction 2), y- (along direction 4), z+ (along direction 5), and z- (along direction 6).

Taking flow in the x+ direction as an example, the boundary conditions are set as follows:

The inflow boundary is located on the yz-plane at x = -128 pixels, with flow in the positive x direction at a velocity.

(Dirichlet boundary condition) (S1a)

(Dirichlet boundary condition) (S1b)

Where V_{in} corresponded to inlet velocity calculated by Reμ/L, in which Re = 0.1, 1, or 10; μ was (Pa s) kinematic viscosity; L was characteristic length calculated by π/s_A, and s_A was the specific surface area.

The outflow boundary is located on the yz-plane at x = 128 pixels, with zero pressure and zero concentration, allowing free outflow based on the computed pressure and concentration gradients.

(Dirichlet boundary condition) (S2a)

(Dirichlet boundary condition) (S2b)

Two pairs of periodic boundaries were set perpendicular to the flow direction. The first pair is on the xz-planes at y = 128 and y = -128 pixels, and the second pair is on the xy-planes at z = 128 and z = -128 pixels. This setup assumes that fluid or substances exiting one boundary re-enter through the opposite boundary in the same state, which simulates continuous flow within the porous structure and reflects the infinite expansibility of the porous medium.

(periodic boundary condition) (S3a)

(periodic boundary condition) (S3b)

(periodic boundary condition) (S3c)

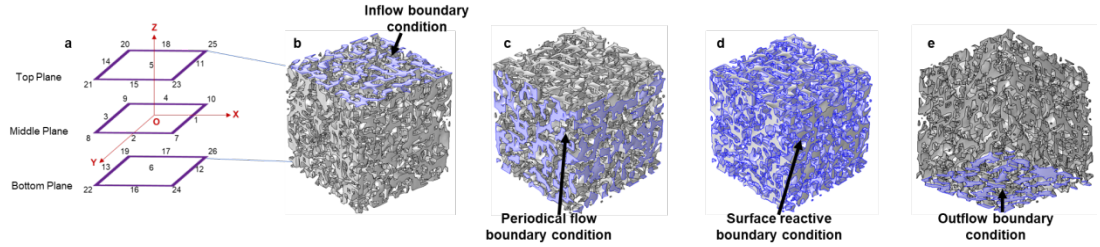
(periodic boundary condition) (S3d)

The remaining boundaries are set as reactive and zero-flux boundaries. According to equation S4a-b, this means that no fluid or substances flow out at these boundaries, and the substances are consumed at a rate calculated by the equation.

(Dirichlet boundary condition) (S4a)

(Neumann boundary condition) (S4b)

Where D (m^2 s^{-1}) corresponded to diffusion coefficient; c was concentration.



Supplementary Fig. 13 | Location of boundary condition: a, Coordinate axes based on the anisotropic growth directions; b, Inflow boundary condition; c, Periodical boundary condition; d, Reactive and zero flux boundary condition; e, Outflow boundary condition.

(2) Wider system in which this block fits

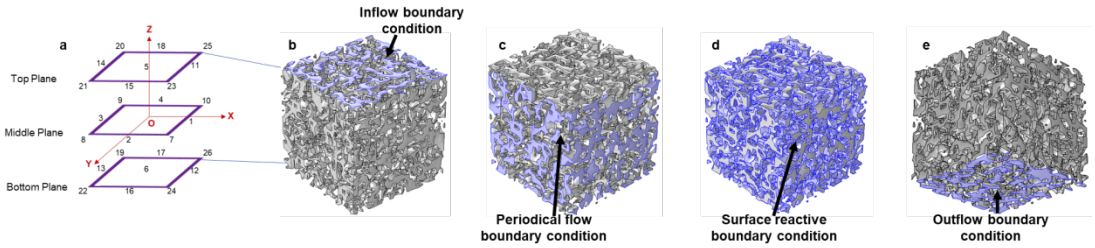
This block fits into a wider system of unidirectional flow being perpendicular to an anisotropic, cube-shaped porous structure, which can be analyzed from six different directions. This represents a fundamental unit of reactive transport process in heterogeneous catalysis, mineral dissolution and environmental transport. We have made corrections on Page 4, Lines 145-150 in the revised manuscript. Added contents are enclosed below.

Page 4, Lines 145-150: To simplify the setup, we focused on cube-shaped porous architecture with unidirectional flow perpendicular to its surfaces. The six faces of the cube were defined as upstream, top, bottom, two sides, and downstream based on the flow direction. This setup resulted in six possible fluid directions: x+ (direction 1), x- (direction 3), y+ (direction 2), y- (direction 4), z+ (direction 5), and z- (direction 6). This represented a fundamental unit of the reactive transport process in heterogeneous catalysis, mineral dissolution, and environmental transport.

Q11: *Supplementary states that the inlet flow velocity is zero, but elsewhere there are 3 different Reynolds numbers.*

Response: The comments are highly appreciated. In original Figure S6D, the "zero flux" was referred to as the "no-flow boundary" on the surface of internal pores, rather than inflow boundary. This leads to misunderstanding because the arrows were placed too close to the inflow boundary. To make clarification, we have redrawn the figure with updated boundary conditions to clearly distinguish between inflow and internal flow boundaries. We have made corrections on Page S19, Lines 211-215 in the revised Supplementary information. Added contents are enclosed below.

Page S19, Lines 211-215:



Supplementary Fig. 13 | Location of boundary condition: a, Coordinate axes based on the anisotropic growth directions; b, Inflow boundary condition; c, Periodical boundary condition; d, Reactive and zero flux boundary condition; e, Outflow boundary condition.

Detailed Responses to Reviewers' Comments

Reviewer #1:

Q1: *A README would help provide some context, but the repository is fairly well organized and it straightforward enough to reproduce the paper as is since the main training script is clear.*

Response: Many thanks for reviewer's positive comment. We have added a README file to provide additional context.

Reviewer #2:

Comments: *The subject of the paper is important, focusing on predicting reaction rates in complex porous media, and challenging, as it considers nontrivial aspects of porous media geometry. In this respect, the paper falls into the category of high-interest and timely subjects that Nature Communications seeks to publish. The previous reviewers raised several concerns about the paper, with the main ones being (1) that the paper is written in a highly specialized, technical style, and that the overall problem setup, approach, and impact are not emphasized, and (2) that the paper does not possess the level of novelty and significance that would justify its publication in Nature Communications. After reading the revised manuscript, it seems to me that these issues have also persisted in the revised version.*

Q1: *The authors attempted to improve the presentation in the new version of the manuscript, but I still find it terse and not accessible to a broader audience. For example, in the introduction, the authors emphasize that the primary subject is the exploration of the "tensorial behavior of reaction rate," yet they never clarify what they mean by this. In physical systems, properties can exhibit tensorial behavior if they link two quantities of a vectorial nature. For example, permeability links flow velocity (a vector) to the pressure gradient (also a vector), forming a tensor. Similarly, the dispersion tensor links the concentration gradient (a vector) with particle flux (another vector). In the introduction, when discussing the tensorial character of flow and transport in porous media, the authors reference papers discussing such quantities. However, in what way is the reaction rate a tensor? In all formulations familiar to me, the reaction rate is a scalar. Although it is locally sensitive to boundary conditions, Peclet and Damköhler numbers, etc., due to its derivation as a solution of a partial differential equation, this does not render it a tensorial quantity. From the context, it becomes clear that what the authors mean by the tensorial character of the reaction rate is that it depends on pore geometry in a more subtle manner than through overall porosity; in particular, pore anisotropy does play a role. In my view, these issues are really important for properly understanding the manuscript and should be expounded upon in the Introduction/Motivation part..*

Response: The comments are highly appreciated. We have replaced "tensorial behavior" with "anisotropy" throughout the manuscript and clarified its meaning following your suggestion on Page 2, Lines 47-50 of the revised manuscript. Added

contents are enclosed below.
"However, local reaction rates depend on pore geometry and flow patterns in a more subtle manner than on overall porosity, particularly in anisotropic structures, which makes it a challenge to understand reactive transport, design catalyst architecture and manipulate catalytic efficiency and selectivity."

Friendly tips: Since we have made over 30 changes in total to address this comment, the revised texts are too long to paste in this document. We kindly invite you to go for details in the places where they are highlighted below in the revised manuscript.

Q2: *Regarding (2): To characterize the anisotropy, the authors link it to the growth probability of the porous medium in 26 directions. This allows them not only to consider artificially created porous media, but also to reconstruct natural porous media in terms of these growth probabilities. While this approach clearly goes beyond a simple characterization of the porous medium by porosity, I believe it misses several important properties of real porous rocks on which the reaction rate depends. First, the pore space of real rocks has a hierarchical, self-similar character, characterized by a well-defined fractal dimension, as evident from the power-law behavior in scattering profiles (see, e.g., Ref. 5 in the manuscript - Zhang et al., 2020, Figs. 4 and 5 there). Second, the connectivity of the pore space, although downplayed by the authors, is tightly linked with the reaction rate (as only connected pores participate in reactions), and this information does not seem to be encoded in the authors' approach. Third, not only the Damköhler number but also the Peclet number will impact the dissolution patterns, and hence the reaction rate. Finally, particularly at large Damköhler numbers, the dissolution of porous media leads to the formation of wormholes (see, e.g., Kalia and Balakotaiah, Modeling and analysis of wormhole formation in reactive dissolution of carbonate rocks, Chemical Engineering Science 62.4 (2007): 919-928), which completely rebuild both pore space and flow paths in the medium, introducing a new, hierarchical structure of connections. A proper characterization of the reaction rate in such a case is a real challenge in the upscaling of reactive transport. Unfortunately, the authors do not take up this challenge, limiting themselves to relatively homogeneous systems without hierarchical channeling structures. All of these factors limit the applicability of the presented method – it should be perfectly suitable to model total reaction rates in porous catalysts, but the application to dissolving rocks is, in my opinion, rather limited.*

Response: Thank you for your suggestion. We have carefully addressed each of the concerns raised.

1. Model scope, limitations, and extension methods

We have clarified the model's scope and limitations, and outlined methods to extend the current workflow to broader applications, in both the Title and Discussion (Page 1, Lines 2-4; Page 13, Lines 432-438) of the revised manuscript. Added contents are enclosed below.

"Visualizing Nexus of Porous Architecture and Reactive Transport in Heterogeneous Catalysis by Deep Learning Computer Vision and Transfer Learning"

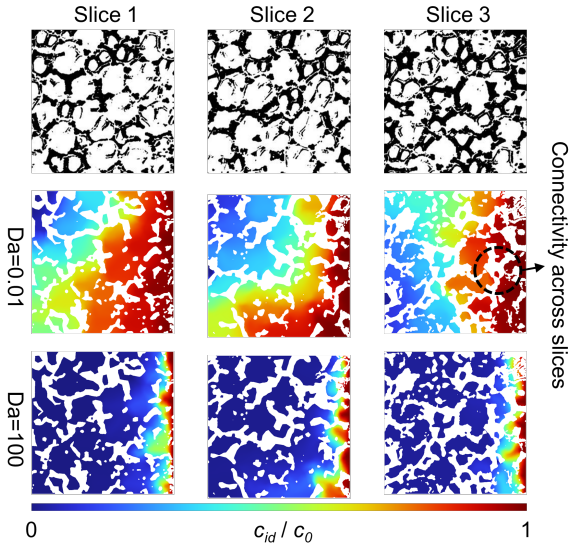
"The current workflow is not well-suited for more variable porous networks (e. g. non-random pores and geotechnical media), especially those with wormholes and hierarchical features, as the source and transfer datasets generated by two-phase QSGS method only account for solid/pore growth on a single scale. To develop these applications, the multi-phase QSGS method (*e. g.* Fast-QSGS61) should be used during dataset construction to capture multiscale pores with self-similar characteristics and complex diagenetic processes."

2. Pore connectivity

To address this issue, we have modified the model to take all slices of the porous structure as input and output the local reaction rates. This allows the model to learn spatial relationships of pores and capture inter-slice connectivity. We have tested the new model's ability to handle the typical connectivity problem on Page S9, Lines 150-154 in the revised Supplementary Information.

"Text S5 The impact of connectivity on model performance

As shown in Figures 3h and S20, some disconnected pores in individual slices were actually connected across slices. The cGAN-TL concentration predictions captured this inter-slice connectivity (Figure S20), demonstrating its ability to detect pore connectivity. This capability stemmed from the cGAN using all slices as input, enabling it to learn the spatial relationships between the pores."



Supplementary Fig. 20 | Porous structure and cGAN-TL concentration predictions showing inter-slice connectivity.

3. Peclet Numbers

There are two forms of the Damköhler number: k/U and kL/D. This study used the kL/D expression, which could be rewritten as $(UL/D)\times(k/U)$, or $\mathrm{Pe}\times(k/U)$. Thus, we did not include Pe as a separate input to avoid unnecessary complexity. We have clarified the meaning of the Damköhler number expressions and their relationship with Pe on Pages 15-16, Lines 520-523 of the revised manuscript.

"$\mathrm{Da}=(UL/D)\times(k_r/U)=\mathrm{Pe}\times(k_r/U)=k_r/(D/L)$ was introduced to represent the ratio of reaction rate to mass transfer rate, where Pe was the Péclet number, U was the flow

velocity, L was the characteristic length calculated by π/s_A, and s_A was the specific surface area[64].”

4. Dissolution and Wormhole Formation

We have removed the dissolving rocks case and emphasized the need to consider multi-process reactive transport in future work in discussion (Pages 13-14, Lines 441-443) of the revised manuscript.

“Future effort will be needed to mitigate these limitations by training of cGAN-FRT that is more suitable for multi-scale, multi-component and multi-process reactive transport across both natural and engineered systems.”

Q3: *Finally, I am also somewhat concerned about the quantity being predicted. R_norm, as defined by Eqs. 12 and 13, is a global quantity, and as such, it is much easier to predict than the local reaction rates at each individual point along the interface. In particular, for large Damköhler numbers, R_norm, by mass conservation, would simply correspond to the total amount of reactant entering the system per unit time (since the aggressive solution does not reach the outlet), and it can be obtained analytically without requiring detailed calculations. At smaller Damköhler numbers, the task is admittedly more challenging, but local reaction rates would seem to be a much more useful quantity to calculate than the overall rate, especially if one is interested in the evolution of the geometry of the reactive system over time.*

Response: As you mentioned, predicting local reaction rates is a more complex and critical challenge. This also represents a limitation for deep learning models like PINNs, which struggle with instability at solid-liquid interfaces. We are pleased to report that we may have found a solution.

A conditional Generative Adversarial Network (cGAN) framework utilizing the Pix2Pix concept has been employed to ensure accurate boundary prediction. Previously, this approach has mainly been leveraged for edge detection and related tasks, such as image colorization and medical image segmentation. This is the first application of the Pix2Pix concept to predict concentration distributions and local reaction rates. As a result, we can convert structure slices into concentration maps with high contour precision, and calculate the local reaction rate as the product of the reaction rate constant and the interface concentration. We have revised the Introduction, Results, Methods, Discussion, and Supporting Information, with key modifications summarized below.

Page 2, Lines 69-79:

“By integrating MSFs with directional and reactive information, it is feasible to establish the structure-property relationship by using multi-model deep learning computer vision (DLCV) techniques[25]. However, conventional DLCV simulation relied on voxel input and exhibited boundary uncertainties[26], making it less suitable for solving solid-liquid interface problems with image input. Owing to predictability, reliability and interpretability, Conditional Generative Adversarial Network (cGAN) excels in pixel-voxel reconstruction and interface prediction tasks, such as 3D object imaging, medical image segmentation and fracture forecasting[27, 28]. These outcomes inspire us to explore the feasibility of developing a mutil-model DLCV based on 3D

reconstrcution-prediction steps to make visual nexus between 2D MSFs and 3D reactive transport. This would be helpful for not only predicting 3D local reaction rate from 2D lateral images, but also providing mechanistic insights into reactive transport of anisotropic porous networks in heterogeneous catalysis."

Page 3, Lines 104-113:

"To predict 3D local reaction rate from 2D images, identify key reaction regions for reactive transport, and overcome geometric diversity limitations in structure-property relationship, we established the transferable cGAN-TL framework with 3D reconstruction-prediction steps and two corresponding generative networks, *i. e*. Generator 1 and Generator 2 (Fig. 1 and Supplementary Fig. 1). In 3D reconstruction step (Generator 1), the lateral images and 3D reconstructed model for porous catalyst served as input and output, respectively. In prediction step (Generator 2), reactive information, inflow direction and slices of reconstructed model were mapped to slices of normalized concentration field, which could then be converted into local reaction rate and normalized effective reaction rate (R_{norm}) that reflected the relative magnitude between effective reaction rate and well-mixed rate (More details in Method section). The reactive information was described by dimensionless Damköhler number (Da) that represented the ratio of solid-liquid reaction rate to solute diffusion rate[34]. The inflow and outflow face corresponded to the left and right sides of the porous catalyst, respectively. To train the generative networks, discriminative network was simultaneously trained for each generative model. Each pair of Generator and Discriminator constituted a cGAN module, as shown in Supplementary Fig. 1."

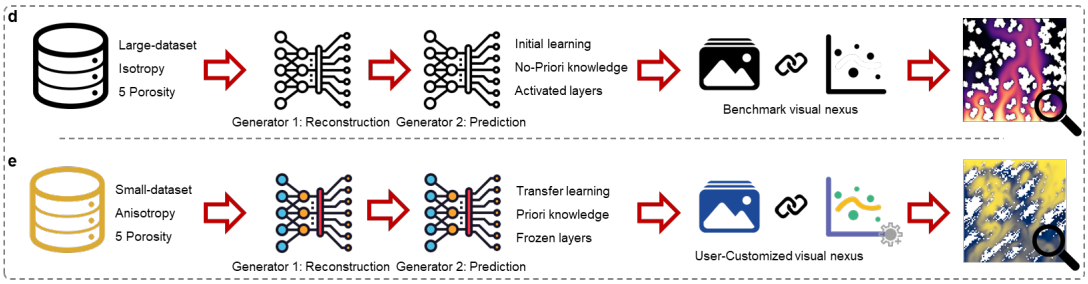


Fig. 1 | 3DCNN-FRT framework description. Workflow for **d,** isotropic architecture and **e,** anisotropic architecture."

Page 5, Lines 164-171:

"Precision of Generator 1 was evaluated by comparing statistical information in original and reconstructed 3D models. Results indicated that the reconstructed model closely matched the original in porosity, connectivity, tortuosity, and coordination number, with errors below 3% and high visual consistency (Figs. 2e-f). For Generator 2, the differences of normalized concentration between actual and predicted slices were plotted in Fig. 2g. Lighter regions indicate deviations, while black represents perfect matches. Remarkably, the deviations were less than 0.1 and were mainly observed near the dead zones. We also plotted the predicted and actual R_{norm} values across the test dataset (Fig. 2h). The cGAN successfully predicted the R_{norm} with 96.3%

accuracy."

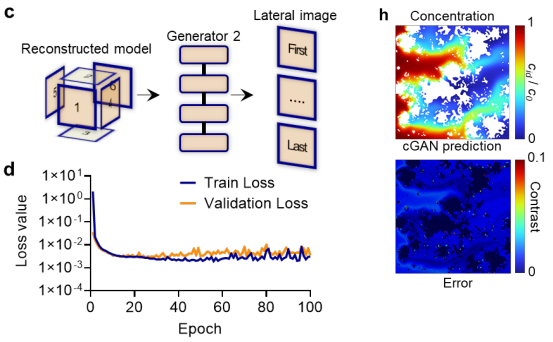


Fig. 2 | Pre-training performance on the source dataset. c, Framework and **d,** Learning curves for Generator 2. **h,** Error in predicted normalized concentration. **i,** Heat map for Generator 2. **PS1:** c_0 represented the initial concentration at inflow side, and c_{id} denoted the dimensionless local concentration. **PS2:** The local reaction rate was calculated as the product of material's reaction rate constant and interface concentration c.

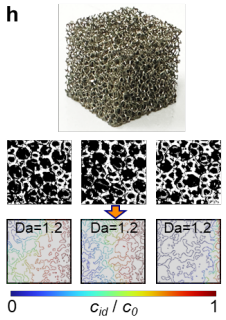


Fig. 3 | Transfer Learning Performance. h, Nickel foam electrode and representative images, along with the corresponding dimensionless concentration at the solid-liquid interface.

Pages 16-17, Lines 552-588:

Model architecture for cGAN-FRT.

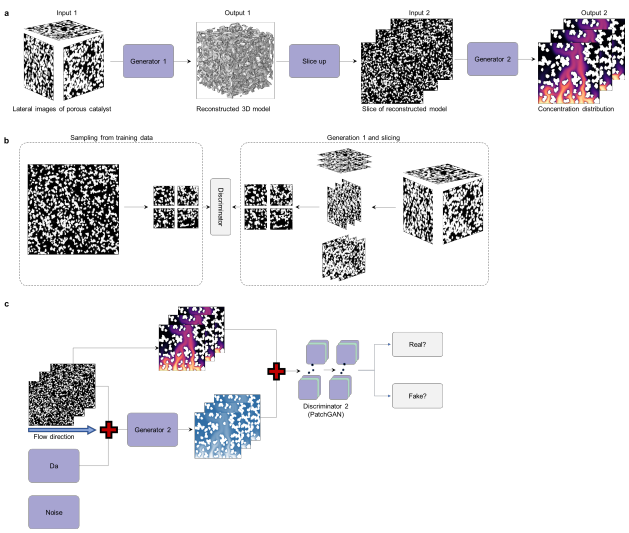
Generator 1 was trained based on SliceGAN[32], which took a latent vector z, reshaped it to 64×4×4×4, and processed it through five 3D transposed convolution layers (kernel size k = 4, stride s = 2, and padding p varying between 2 and 3) to iteratively upsample the feature maps. The final output was a 3×64×64×64 volume, followed by a softmax activation. For discrimination 1, a 2D network (memory size ~11 MB) received 3-channel 64×64 inputs and processed them through five 2D convolution layers (kernel size k = 4, stride s = 2, and padding p varying between 1 and 0). The discriminator's output was a 1×1×1 scalar, representing the real/fake classification. Further details about Generator 1 can be found in SliceGAN[32].

Generator 2 was designed using a U-Net-based architecture. The model took a 3×D×H×W input volume (*e. g.* 3×64×64×64) and a scalar value input (Da), which was embedded into a feature vector through a fully connected network. The embedded Da feature was concatenated with the input volume and processed through a series of

five downsampling layers (kernel size k = 4, stride s = 2, padding p = 1 or 2). These layers progressively reduced the spatial dimensions of the input while extracting discriminative features. The downsampled features were then passed through five upsampling layers (kernel size k = 4, stride s = 2, padding p = 1), with skip connections from corresponding downsampling layers to restore spatial resolution and improve output quality. The final output was a 3×D×H×W volume, produced by a transposed convolution layer followed by a Tanh activation function to scale the output within the range [-1, 1].

The Discriminator 2 was built to distinguish between real and fake 3D images using both image data and an embedded force feature. It took a grayscale image, a scalar Da input, and a color image as inputs. The Da input was first processed through the Encoder to produce a feature vector, which was then reshaped and broadcasted to match the spatial dimensions of the input images. These features were concatenated with the grayscale and color images along the channel dimension, forming a combined input. The model then passed this concatenated input through a series of four convolutional layers with LeakyReLU activation and batch normalization (for the second and third layers). The final layer reduced the output to a single scalar, representing the real/fake classification. The model employed convolutional layers with kernel size k = 4, stride s = 2 (except the last layer with stride s = 1), and padding p = 1 or 0, progressively extracting discriminative features from the input.

In Generator 1, the input was three lateral images from the QSGS porous model, and the output was a reconstructed 3D model. The discriminator evaluated quality by comparing slices of the reconstructed model from different angles with the input images[27]. In Generator 2, the Da number served as reactive information, and structural information came from slices of the reconstructed model in Generator 1, spaced one pixel apart. The output was slices of dimensionless concentration (c_{id}) field. All slices for both input and output were cut along the flow direction. According to analysis for R_{norm} data structure shown in Supplementary Text 4 and Fig. 15, we fine-tuned the source models by using feature representation transfer learning (FRT)[48].



Supplementary Fig. 1 | Model architecture for cGAN-FRT. (a) two-steps workflow. (b) Architecture for Generator 1. c, Architecture for Generator 2.

Friendly tips: Since we have modified more than 2300 words totally to address this comment, the revised texts are too long to paste in this document. We kindly invite you to go for details in the places where they are highlighted below in the revised manuscript.

Detailed Responses to Reviewers' Comments

Reviewer #3:

Comments: *I've read with interest the new version of the manuscript, which in my opinion is definitely better than the original version, primarily because the authors were able to recover local reaction rates. This brings the level of novelty and significance expected for Nature Communications articles. Most of my previous comments were addressed satisfactorily. The remaining points of disagreement are discussed below (authors' responses from their rebuttal letter are in italic):*

Q1: *Wormholes arise spontaneously due to dissolution-driven self-organization in initially random pore architectures. Thus, the phrase "non-random pores" is overly simplistic and potentially misleading. I suggest making this discussion more nuanced. In particular, it should be stated clearly that the model, in its current form, is not capable of reproducing the evolution of porous media in the wormholing regime.*

Response: The comments are highly appreciated. We have removed the misleading phrase of "non-random pores", and clarified that our current model cannot reproduce wormholes on Page 13, Lines 443-446 of the revised manuscript. Added contents are enclosed below.

"However, the current workflow relies on two-phase QSGS datasets, which only captured solid–pore growth at a single scale, but was incapable of reproducing the multiscale, self-organizing evolution of geological porous media, especially in the wormholing regimes[50-52]. To develop these applications, the multi-phase QSGS method (*e. g.* Fast-QSGS[53]) should be used during dataset construction to capture multiscale pores with self-similar characteristics and complex diagenetic processes."

References

50 Zhang, R., Liu, S., He, L., Blach, T. P. & Wang, Y. Characterizing Anisotropic Pore Structure and Its Impact on Gas Storage and Transport in Coalbed Methane and Shale Gas Reservoirs. *Energy & Fuels* **34**, 3161-3172, doi:10.1021/acs.energyfuels.0c00109 (2020).

51 GOLFIER, F., ZARCONE, C. BAZIN, B. LENORMAND, R. LASSEUX, D. & QUINTARD, M. On the ability of a Darcy-scale model to capture wormhole formation during the dissolution of a porous medium. *Journal of Fluid Mechanics* **457**, 213-254, doi:10.1017/S0022112002007735 (2002).

52 Wang, M., Wang, J., Pan, N. & Chen, S. Mesoscopic predictions of the effective thermal conductivity for microscale random porous media. *Physical Review E* **75**, 036702, doi:10.1103/PhysRevE.75.036702 (2007).

Q2: *I cannot agree with this reasoning. As is well-established in the literature, dissolution patterns depend on two independent dimensionless numbers: the Peclet and Damköhler numbers - see, for example, Golfier et al., "On the ability of a Darcy-scale model to capture wormhole formation during the dissolution of a porous medium." Journal of Fluid Mechanics 457 (2002): 213–254, especially Fig. 12. As shown there, the morphological phase space is twodimensional and cannot be*

reduced to a single parameter. Fundamentally, the patterns depend on three physical parameters – the reaction rate (k), diffusion coefficient (D), and fluid velocity (U) – from which two independent dimensionless numbers can be constructed; this is not "unnecessary complexity" but a reflection of the underlying physics.. While it is true that Da = Pe × (k/U), expressing it this way eliminates the explicit dependence on U. The authors should therefore revise this part of the discussion accordingly, but also explicitly address the following question: how is it that their dissolution patterns depend only on the Damköhler number?

Response: Agree and corrected accordingly. Our earlier models generated the results under three Re values and applied transfer learning to recalibrate system parameters (*e. g*. diffusion coefficient and structural anisotropy), indicating that reactive transport was not governed solely by Da value. Following your suggestions, we have rebuilt the models by introducing Pe as an input, and removed the previously unsupported reasoning from the revised manuscript. Modified contents are enclosed below.

1. We added detailed discussion on the models as function of physical parameters. Pages 4-5, Lines 151-158 of the revised manuscript.

"Each model and its dataset represented a system with specific solute–solvent pair and fixed structural anisotropy. Each system included two operating variables (Pe and Da) and four constant parameters (anisotropy factor g_i, viscosity μ, fluid density ρ, and diffusion coefficient D). The operating variables served as model inputs, while the constant parameters defined the transfer-learning domains. Any change of the constant parameters required retraining the model through transfer learning."

"Since Reynolds number (Re=Pe×D×ρ/μ) was linearly proportional to the Péclet number (Pe) within a given dataset, only Pe was used as the hydrodynamic descriptor."

2. We rebuilt the models to accept three inputs (structure images, Pe, and Da) and produce one output (corresponding concentration field), from which the local reaction rate could be calculated. Page 16, Lines 557-579 of the revised manuscript.

"Generator 2 was designed with a 3D U-Net architecture. The model accepted a 1×D×H×W grayscale input volume (*i. e*. 1×64×64×64) together with two scalar inputs (Da and Pe). These scalars were first embedded by a two-layer fully connected network into a 64-dimensional feature vector. The embedded feature map was then broadcast across the depth, height, and width dimensions and concatenated with the grayscale volume, yielding a 65-channel tensor. This tensor passed through five 3-D down-sampling blocks (kernel k=4, stride s=2, padding p=1; the first block omitted BatchNorm). The encoder outputs were fed into five 3D up-sampling blocks (kernel k=4, stride s=2, padding p=1), with skip connections from the corresponding encoder layers restoring spatial detail. Last, a 3D transposed-convolution layer followed by a Tanh activation produced a 3×D×H×W RGB volume whose voxel intensities lied in [-1, 1].

Discriminator 2 was built to distinguish between real and fake 3D images by incorporating the embedded Pe and Da features. It received a grayscale volume, scalar

Da and Pe values, and a color volume as inputs. The scalar inputs were first processed by an encoder to produce a 64-dimensional feature vector, which was then reshaped and broadcast to match the spatial dimensions of the input volumes. These broadcast features were concatenated with the grayscale and color volumes along the channel dimension, forming a 68-channel tensor. The combined input was then passed through a sequence of three 3D convolutional layers (kernel k=4, stride s=2, padding p=1), each being followed by a LeakyReLU activation. Batch normalisation was applied to the second and third layers. A final 3D convolution (kernel k=4, stride s=1, padding p=1) produced a PatchGAN grid of logits, with each element evaluating the authenticity of its local 4×4×4 patch. Binary-cross-entropy with logits was used as the adversarial loss, enabling the network to progressively extract discriminative volumetric features being sensitive to the conditioning parameters."

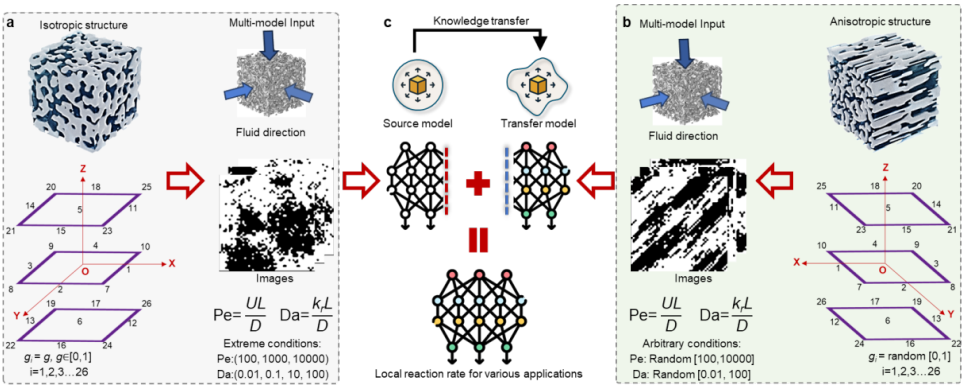
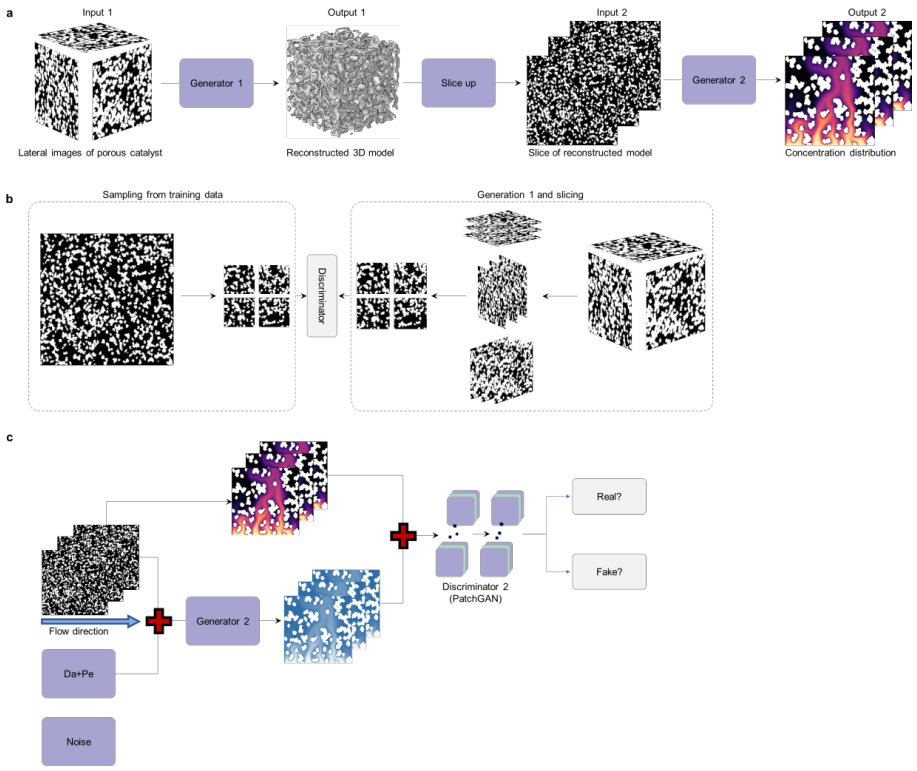


Fig. 1 | 3DCNN-FRT framework description. Input for a, isotropic architecture and b, anisotropic architecture. c, Transfer learning strategy.



Supplementary Fig. 1 | Model architecture for cGAN-FRT. a, two-steps workflow. b, Architecture for Generator 1. c, Architecture for Generator 2.

3. The dataset construction was also revised. Page 13, Lines 457-458 and Page 14, Lines 472-474 of the revised manuscript.

"Input data for Generator 2 included Damköhler number (Da), Péclet number (Pe) and structure slices."

"Each structure had six inlet directions, five Da values (*i. e.* 0.01, 0.1, 1, 10, 100) and three Pe values being typical of heterogeneous catalysis (*i. e.* 100, 1000, 10000), and this resulted in the construction of 14400 scenarios (4×4×10×6×5×3)."

4. We removed the unsupported reasoning and updated the corresponding discussion. Page 5, Lines 175-177 of the revised manuscript.

"For Generator 2, the predicted and actual R_{norm} were plotted with respect to Da and Pe values (Fig. 2g). The cGAN not only predicted R_{norm} with 97.3% accuracy, but also effectively captured the dominant role of Pe in R_{norm} variation under high Da value."

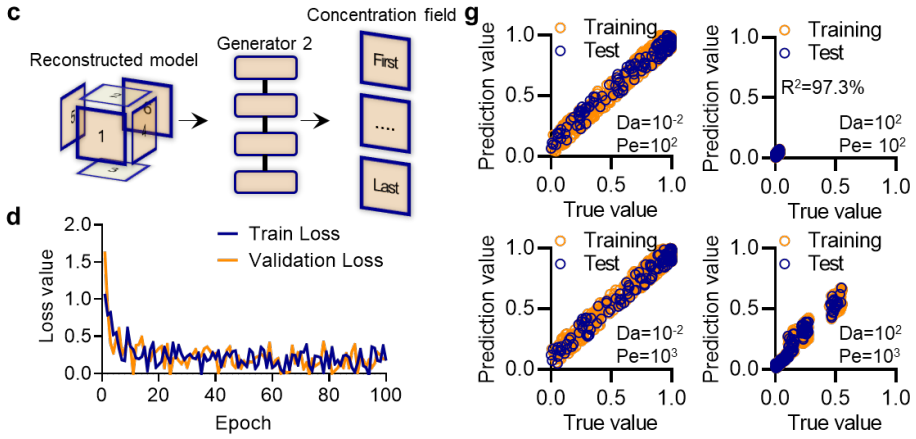


Fig. 2 | Pre-training performance on the source dataset. c, Framework and d, Learning curves for Discriminator 2. g, Predicted results and error for the dimensionless concentration field.

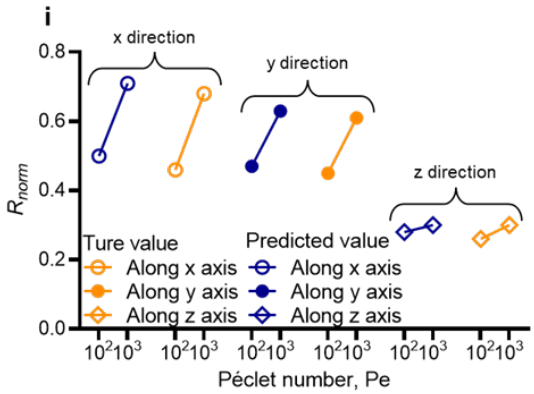


Fig. 3 | Transfer Learning Performance. i, Comparative analysis between tested and predicted R_{norm} under different Re and fluid directions.

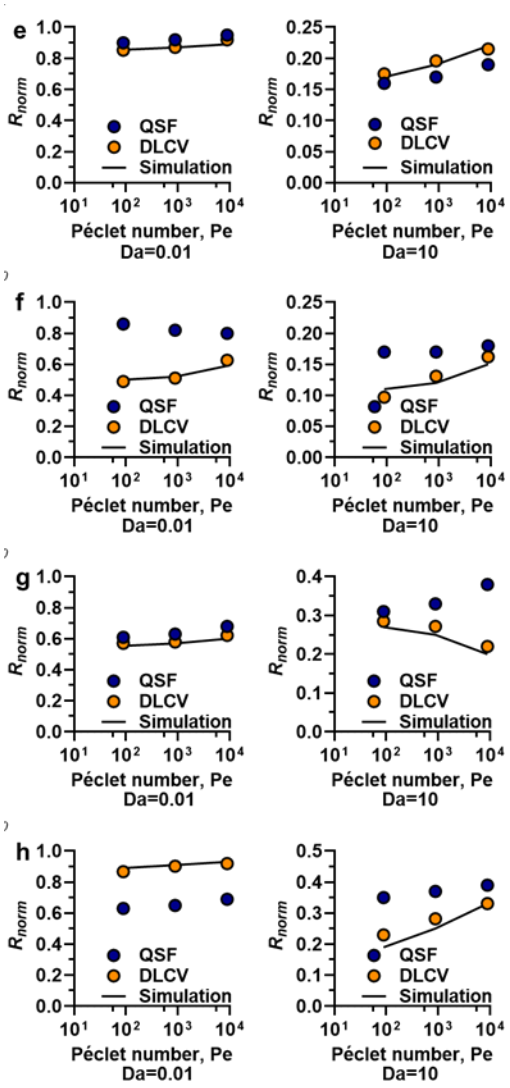


Fig. 4 | Accuracy and convenience comparison for different methods. e, Concentration distribution and f, Tensorial R_{norm} variations for axis anisotropic architecture. g, Concentration distribution and h, Tensorial R_{norm} variations for diagonal anisotropic architecture.

Friendly tips: Since we have introduced more than twenty modifications to address this comment, the revised texts are too long to paste in this document. We kindly invite you to go for details in the places where they are highlighted below in the revised manuscript. Because the only dataset modification is the conversion of output Re into input Pe (while Generator 1, porous structures and concentration field remain unchanged), the significance analysis in Figure 5 remains largely the same.

Detailed Responses to Reviewers' Comments

Reviewer #3:

Comments: *I am generally happy with the revisions the authors have made. Perhaps one minor comment: in the new sentence — "However, the current workflow relies on two-phase QSGS datasets, which only captured solid–pore growth at a single scale, but was incapable of reproducing the multiscale, self-organizing evolution of geological porous media, especially in the wormholing regimes [50-52]" — I do not see how references [50] and [52] relate directly to wormholing or multiscale evolution. I suggest keeping there only Ref. 51. Except for this, I believe the paper is ready for publication.*

Response: Many thanks for reviewer's positive comment. We have made modifications accordingly on Page 13, Line 446 in the revised manuscript.

However, the current workflow relies on two-phase QSGS datasets, which only captures the solid–pore growth at a single scale, but is incapable of reproducing the multiscale, self-organizing evolution of geological porous media, especially in the wormholing regimes[50].

References:

50 Golfier, F., Zarcone, C. Bazin, B. Lenormand, R. Lasseux, D. & Quintard, M. On the ability of a Darcy-scale model to capture wormhole formation during the dissolution of a porous medium. *Journal of Fluid Mechanics* **457**, 213-254, doi:10.1017/S0022112002007735 (2002).