Knowledge-Guided Transfer Learning for Modeling Subsurface Phenomena Under Data Paucity

Nikhil Muralidhar
Department of Computer Science
Stevens Institute of Technology
Hoboken, NJ 07030
nmurali1@stevens.edu

Nicholas Lubbers
Los Alamos National Labs
Los Alamos, NM 87545
nlubbers@lanl.gov

Mohamed Mehana
Los Alamos National Labs
Los Alamos, NM 87545
mzm@lanl.gov

Naren Ramakrishnan
Department of Computer Science
Virginia Tech Arlington, VA 22203
naren@cs.vt.edu

Anuj Karpatne
Department of Computer Science
Virginia Tech Blacksburg, VA 24061
karpatne@vt.edu

Abstract

Knowledge transfer from machine learning (ML) models, pre-trained on large corpuses has been leveraged effectively in domains like natural language processing and computer vision to improve model generalization. The knowledge transfer prowess of ML and especially deep learning (DL) models has been demonstrated to be especially effective under data paucity of the target task. Many scientific phenomena require the execution of costly simulations to estimate the process of interest. Predicting molecular configuration of fluids confined in porous media is one such problem of extreme relevance in many scientific applications, the study of which requires the execution of expensive Molecular Dynamics (MD) simulations. However, due to the cost of MD, large scale simulations become intractable. Hence, in this work, we develop a novel science-guided deep learning framework NanoNet-SG to emulate MD simulations. Our proposed NanoNet-SG model leverages scientific domain knowledge in conjunction with knowledge from pre-trained knowledge bases for estimating molecular configuration of fluid mixtures. Through rigorous experimentation, we demonstrate that our proposed NanoNet-SG model yields good generalization performance improvements (i.e., overall performance improvement of greater than **16**% over baselines) and yields predictions that are consistent with known scientific domain rules despite being trained on a low volume of MD simulation data (i.e., data paucity).

1 Introduction

An especially popular use of ML models (primarily deep learning models) is to function as relatively inexpensive *surrogates* to emulate the behavior of costly scientific processes. Owing to their ability to model complex functions, various ML (surrogate) models have been proposed for modeling fluid flows Brunton et al. [2019], Wang et al. [2020], Raissi et al. [2019].

Deep learning models, which are usually capable of effectively modeling complex functions, are plagued by their inability to learn good representations under data paucity. However, certain scientific modeling contexts inherently suffer from data paucity owing to expensive simulation and data collection costs. One such scientific application, which is the focus of this work, is modeling fluid properties in nano-porous media Amini and Mohaghegh [2019], Monteiro et al. [2012]. Fluid properties in nano-porous media are important in applications like petroleum engineering Amini and

NeurIPS 2022 AI for Science Workshop.

Mohaghegh [2019], Lubbers et al. [2020], Santos et al. [2020], carbon sequestration Li et al. [2020a] and energy conservation. With the increasing energy demands and technological advances like hydraulic fracking, unconventional resources like shale formations are being leveraged as potential sources of energy. Hydrocarbons like methane and ethane which are abundant in nano-pores in large shale formations, serve as feasible and effective candidates to address increasing energy demands. One of the main challenges to modeling fluid properties (e.g., viscosity, density, adsorption patterns) in shale reservoirs is that the size of the pores comprising the hydrocarbons of interest are often less than 50 nanometers (nm) in width. It has been observed that fluid properties in such small pores (i.e., under nano-confinement) significantly deviate from their properties under bulk conditions, i.e. in larger spaces where interactions with the surface are rare. Thus, modeling of fluid properties under nano-confinement cannot be faithfully represented using classical fluid models (e.g. Navier-Stokes Equations), and so often researchers employ more expensive molecular dynamics (MD) simulations. Owing to the high computational demands of modeling fluid properties via MD, it is infeasible to compute fluid properties directly for all possible pore spaces.

In this work we develop NanoNet-SG , a novel science-guided framework to emulate MD simulations for fluid mixtures. Specifically, to alleviate the cost of expensive MD simulations of fluid mixtures, NanoNet-SG leverages limited fluid mixture simulation data and relies mostly on knowledge gained from relatively simpler (and cheaper) MD simulations of singular fluids in related experimental contexts. This knowledge from modeling singular fluid simulations (with relatively large data volumes) to modeling fluid mixtures (problem of interest) with low data volumes is performed via *Science-Guided Knowledge Transfer* in our proposed NanoNet-SG model.

Our contributions are as follows:

• We develop a novel and light-weight *inductive transfer learning* framework NanoNet-SG for modeling the density of fluid mixtures under nano-confinement.

• Our NanoNet-SG model leverages scientific domain knowledge to influence training dynamics.

• We demonstrate through rigorous experimentation, the generalization ability of NanoNet-SG .

• We also verify the improved domain consistency of NanoNet-SG relative to variants without science guidance.

2 Background

To counter the effects of modeling under data paucity, we leverage the well studied transfer learning paradigm to develop our NanoNet-SG model.

2.1 Transfer Learning

Essentially, the paradigm of transfer learning (effectively used in various natural language Ruder et al. [2019], computer vision Li et al. [2020b] and healthcare Gupta et al. [2020], Chen et al. [2019]) relaxes the fundamental assumption in traditional ML paradigms of distributional similarity in training and testing domains. Instead, transfer learning approaches focus on developing ML frameworks with the ability to apply knowledge and skills learned on previous tasks to a related but novel task wherein data might not necessarily share distributional similarity with previously encountered tasks. Transfer learning is especially effective in training ML models on tasks with low volumes of data, by first training the model on (i.e., leveraging knowledge from) a related task with relatively larger data volumes. We now re-state relevant preliminaries detailed in Pan and Yang [2009] specific to transfer learning.

Let us assume a *domain* $\mathcal{D}$ comprising a feature space $\mathcal{X}$ and a marginal distribution $\mathrm{P}(\mathbf{x})$ where $\mathbf{x} \in \mathcal{X}$ is a particular data instance. In a specific domain $\mathcal{D} = \{\mathcal{X}, \mathrm{P}(\mathbf{x})\}$, let us define a *task* $\mathcal{T}$ to contain a label space $\mathcal{Y}$ and a prediction function $\mathrm{f}(\cdot)$ to be learned from training data. For an instance $\mathbf{x}^i \in \mathcal{X}$, $\mathrm{f}(\mathbf{x}^i)$ yields the corresponding estimate $y^i \in \mathcal{Y}$.

Consider a *source* domain $\mathcal{D}_S$ (and source learning task $\mathcal{T}_\mathrm{S}$) and a *target* domain $\mathcal{D}_T$ (target learning task $\mathcal{T}_\mathrm{T}$). Let $\mathrm{C_S} = \{(\mathbf{x}_S^1, y_S^1), .., (\mathbf{x}_S^n, y_S^n)\}$ and $\mathrm{C_T} = \{(\mathbf{x}_T^1, y_T^1), .., (\mathbf{x}_T^m, y_T^m)\}$ represent the training data corpus in the source and target domains, respectively, such that $|\mathrm{C_T}| << |\mathrm{C_S}|$, where $|\cdot|$ indicates cardinality of the data corpus (i.e., available training data volume).

Definition 1 *Given a source domain and task* $(\mathcal{D}_{\mathrm{S}}, \mathcal{T}_{\mathrm{S}})$ *and a related but distinct target domain and task* $(\mathcal{D}_T, \mathcal{T}_T)$ *such that* $|\mathrm{C_T}| << |\mathrm{C_S}|$*, transfer learning aims to improve the learned predictive function* $\mathrm{f}_T(\mathrm{C_T})$ *by leveraging knowledge from* $(\mathcal{D}_{\mathrm{S}}, \mathcal{T}_{\mathrm{S}})$.

Following from Def. 1, a setting in which $\mathcal{T}_{\mathrm{S}} \neq \mathcal{T}_{\mathrm{T}}$ is specifically characterized as *Inductive Transfer Learning* (ITL) Pan and Yang [2009].

2.2 Fluid Constitution Under Nano-Confinement

The overall structure of a *density profile* of a fluid confined in a nano-pore comprises of two regions.

Peak Region. The density distribution of the fluid closest to the pore walls which exhibit increased concentration (i.e., *peaks*) due to adsorption onto the pore walls. **Bulk Region**. Intuitively, the bulk region corresponds to any part of the density profile *away* from the walls where the adsorption effects due to the pore walls is negligible. For simplicity, we assume that the bulk region corresponds to the density distribution in and near the center of the pore.

3 Problem Formulation

Setup. We propose a novel ML emulator that leverages the power of inductive transfer learning (ITL) and scientific domain knowledge to develop a novel science-guided architecture (NanoNet-SG) for fluid density estimation of a mixture M of fluids under nano-confinement.

Large scale MD simulations for fluid mixtures are expensive to setup and execute, leading to data paucity in the target domain $\mathcal{D}_{\mathrm{T}}$. To alleviate the effect of data paucity in $\mathcal{D}_{\mathrm{T}}$, we leverage ITL and pre-train models to encode knowledge from the relatively simpler context of single fluid nano-confinement for each of the constituent fluids in mixture M. We now detail the setting of interest.

3.1 Target Domain Description

Our problem comprises of a target domain $\mathcal{D}_{\mathrm{T}}$, target task $\mathcal{T}_{\mathrm{T}}$ and function $f_T : \mathbb{R}^{In} \rightarrow \mathbb{R}^{|M| \times l}$ that estimates fluid density profiles for a mixture of 'k' constituent fluids $M = \{F_1, F_2, .., F_k\}$. $f_{\mathrm{T}}(\cdot)$ is trained on a limited data corpus $\mathrm{C_T} = \{(\mathbf{x}_T^1, \mathbf{Y}_T^1), ..., (\mathbf{x}_T^m, \mathbf{Y}_T^m)\}$ where $\mathbf{x}_T^i$ corresponds to the i^{th} set of input features and $\mathbf{Y}_T^i \in \mathbb{R}^{|M| \times l}$ (where l is the estimation horizon of the density profiles) is the corresponding set of target density profiles.

3.1.1 Target Task Feature Description.

Each set of input features $\mathbf{x}_T^i = \{\mathbf{e}^i, \boldsymbol{\upsilon}^i, \hat{\mathbf{y}}_{S_1}^i, \hat{\mathbf{y}}_{S_2}^i, .., \hat{\mathbf{y}}_{S_k}^i\}$ consists of three components (i) the experimental setting $\mathbf{e}^i$ (ii) the spatial pore encoding $\boldsymbol{\upsilon}^i$ and (iii) the set of predictions $\hat{\mathbf{y}}_{S_j}^i$ from pre-trained single fluid models in the source domain $\mathcal{D}_{S_j}$ for each constituent fluid $F_j \in M$.

Experimental Setting. In $\mathbf{e}^i = \{\mathrm{w}^i, \mathrm{t}^i, \mathrm{p}^i, \psi_1^i, .., \psi_k^i\}$. $\mathrm{w}^i \in \mathbb{R}^{1\times1}$ indicates the width of the pore being modeled, where the mixture of fluids is confined (measured in Angstroms), $\mathrm{t}^i \in \mathbb{R}^{1\times1}$ signifies the temperature in the pore (measured in Kelvins) and $\mathrm{p}^i \in \mathbb{R}^{1\times1}$ the total pressure within the pore (measured in Atmospheres). Each $\psi_j^i \in \mathbb{R}^{1\times1}$ value indicates the *partial pressure* for each constituent fluid $F_j \in M$ such that $\mathrm{p}^i = \sum_{j=1}^{|M|} \psi_j^i$.

Spatial Pore Encoding. Each spatial pore encoding $\boldsymbol{\upsilon}^i \in \mathbb{R}^{d\times l}$ comprises 'd' different spatial encodings of every point in the pore at which the density profile is to be estimated (called the estimation horizon of size 'l')[1] for each constituent fluid. Note that $\boldsymbol{\upsilon}^i$ is the same for all fluids in the mixture for a particular instance 'i'. In our experiments $d = 2$, consisting of two separate encodings for every point in the estimation horizon. First encoding being a binary representation indicating if each point in the estimation horizon is internal to the pore or outside (i.e. pore wall). The second encoding is a continuous representation of the *inverse distance* of each point in the pore from the nearest pore wall.

[1] we employ a uniform estimation horizon governed by the largest pore width in our corpus. Spatial encodings for smaller pores are 'periodically padded' to achieve uniform lengths of spatial encodings for all pore sizes.

Pre-Trained Density Estimates. $\hat{\mathbf{y}}^i_{S_j} \in \mathbb{R}^{1\times l}$ represents the prediction (by pre-trained models in the source domain) of fluid density for fluid F_j at each point in the estimation horizon for experimental setting $\mathbf{e}^i$ and spatial encoding $\boldsymbol{\upsilon}^i$.

3.2 Source Domain Description

Owing to the data paucity due to simulation cost and complexity of MD simulations in the target domain $\mathcal{D}_\mathrm{T}$, we leverage knowledge from a related (albeit simpler) paradigm of modeling fluid density under nano-confinement of singular fluids. For each fluid $F_j \in M$, our source context is $(\mathcal{D}_{S_j}, \mathcal{T}_{S_j}, f_{S_j} : \mathbb{R}^{In_S} \rightarrow \mathbb{R}^{Out_S})$ where $\mathcal{D}_{S_j}$ is the source domain comprising of the training data $\mathrm{C}_{S_j} = \{(\mathbf{x}^1_{S_j}, \mathbf{y}^1_{S_j}), .., (\mathbf{x}^n_{S_j}, \mathbf{y}^n_{S_j})\}$ used to train a learning model to estimate function f_{S_j} for the single fluid density estimation task $\mathcal{T}_{S_j}$. $\mathbf{x}^i_{S_j} = \{\mathbf{e}^i, \boldsymbol{\upsilon}^i\}$ and $\mathbf{y}^i_{S_j} \in \mathbb{R}^{1\times l}$. The experimental setting for the source domain comprises of $\mathbf{e}^i = \{\mathrm{w}^i, \mathrm{t}^i, \mathrm{p}^i\}$ without the partial pressures as the setting comprises only of a single fluid. $\boldsymbol{\upsilon}^i \in \mathbb{R}^{d\times l}$ is encoded same as in the target domain. The set of tasks $T_S = \{T_{S_1}, .., T_{S_{|M|}}\}$ results in $|M|$ pre-trained models, each trained to estimate fluid density for a particular fluid $F_j \in M$.

3.3 Domain Adaptation with NanoNet

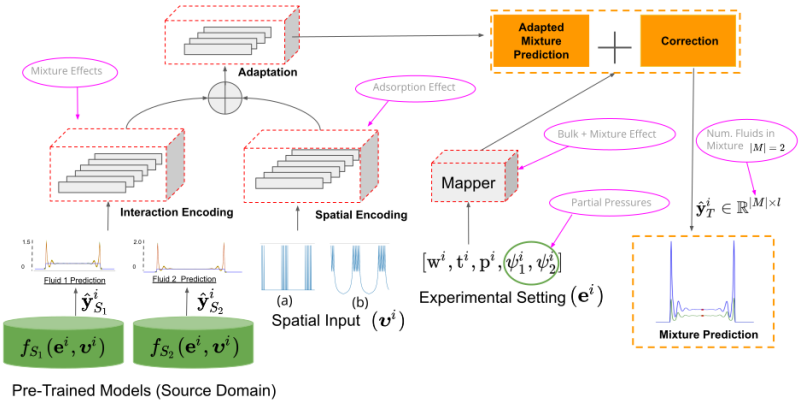


Figure 1: NanoNet-SG Architecture for Modeling Two-Fluid Mixture with only ∼5,100 Learnable Parameters.

Fig. 1 outlines the inductive transfer learning architecture of our proposed NanoNet-SG model. The model architecture comprises of four sub-modules (outlined as dotted-red boxes). The **Interaction Encoding** module accepts predictions from the pre-trained single fluid models ($f_{S_i}(\cdot)$) and learns to model the interaction and *mixture effects* relevant to the target context. The **Spatial Encoding** module captures the relevant spatial information from the $\boldsymbol{\upsilon}^i$ spatial input of the current pore context. The **Adaptation** module combines outputs of the *interaction*, *spatial encoding* to yield the *Adapted Mixture Prediction*. Finally, the **Mapper** module which is conditioned on experimental setting $\mathbf{e}^i$, also comprising partial pressures of each constituent fluid in the mixture being modeled produces outputs used to *correct* the *Adapted Mixture Predictions* to yield the final output of density profile predictions. Finally, the *adapted mixture prediction* is combined with appropriate *corrections* to yield the final predicted output of the fluid density for each constituent fluid in the mixture M. The *Interaction Encoding, Spatial Encoding* & *Adaptation* modules leverage 1×1 convolutions Lin et al. [2013] while the *Mapper* module is a set of fully connected blocks.

3.4 Science-Guided Domain Adaptation with NanoNet-SG

The overall NanoNet-SG training loss is a combination of the data-driven and the science guided loss functions and is represented in Eq. 1[2].

[2] $\lambda = 1e^6$ by manual tuning on training data.

$$\mathcal{L} = \mathcal{L}_{\text{Data}} + \lambda \mathcal{L}_{SG} \tag{1}$$

Data Loss ($\mathcal{L}_{Data}$). The NanoNet-SG model is trained on data corpus C_T with data-driven mean-squared error loss (Eq. 2). Here, $\mathbf{y}_T^{ij} \in \mathbb{R}^{1\times l}$ represents the density profile of the ith instance for fluid $F_j \in M$, and $\hat{\mathbf{y}}_T^{ij} \in \mathbb{R}^{1\times l}$ represents the corresponding density profile prediction made by NanoNet-SG model for input $\mathbf{x}_T^i$.

$$\mathcal{L}_{\text{Data}} = \sum_{j=1}^{|M|} L_{\text{Data}}^j \qquad L_{\text{Data}}^j = \frac{1}{l * |\text{C}_\text{T}|} \sum_{i=1}^{|\text{C}_\text{T}|} ||\hat{\mathbf{y}}_T^{ij} - \mathbf{y}_T^{ij}||_2^2 \tag{2}$$

Science-Guided Loss ($\mathcal{L}_{SG}$). In order to improve model generalization and produce domain consistent representations under data paucity, we augment the NanoNet-SG training pipeline with science-guidance derived from theoretical knowledge about fluids under nano-confinement. Specifically, we adopt the following well understood and accepted phenomenon.

Remark 1 We draw from thermodynamics that total density 'ξ' of a fluid under nano-confinement is monotonically non-decreasing with increase in pressure (p) for a given pore width (w), temperature (t).

Let ξ^{ij} represent the total density, calculated as the area under a predicted density curve $\hat{\mathbf{y}}_T^{ij}$. Further, without loss of generality, we represent ξ_p^{ij} to be the total density for an experimental setting $\mathbf{e}^i$ with pressure p.

The monotonically non-decreasing relationship between ξ and pressure is incorporated directly into the training loss to yield the novel NanoNet-SG model as defined in Eq. 3.

$$\begin{aligned} L_{SG}^j &= \sum_{i=1}^{|\text{C}_\text{T}|} g(\xi_p^{ij}, \xi_{p+\Delta p}^{ij}) + g(\xi_{p-\Delta p}^{ij}, \xi_p^{ij}) \qquad g(a,b) = \begin{cases} 0 & a \le b \\ b-a & \text{otherwise} \end{cases} \\ \mathcal{L}_{SG} &= \sum_{j=1}^{|M|} L_{SG}^j \end{aligned} \tag{3}$$

We choose the step-size Δ resulting from uniformly dividing our entire domain of interest (i.e., pressure range of interest as described in Table. 1) into 100 bins.

4 Experimental Setup

4.1 Dataset Description

As described in Section. 3, we employ two separate types of datasets. The first type (i.e., the target domain) being the molecular dynamics simulations of fluid mixtures (i.e., a 2 fluid mixture of methane and ethane) and the source domain(s) being data corpuses from MD simulations of each of the constituent fluids (i.e., separate datasets of methane and ethane simulated in isolation). The pressure and temperature ranges as well as the size of each dataset are described in Table 1. Our experimental setups comprise distinctly of *fluid mixture* and *single fluid* modeling tasks. In this work, we focus on modeling the two hydrocarbons *methane* and *ethane* both in the single fluid and mixture settings. The experimental setup ranges and data generation processes are primarily governed by domain experts in petroleum engineering. Although we focus on the task of fluid *density* estimation, the proposed NanoNet-SG model can be easily adapted to estimate other properties like *viscosity*.

4.2 Baseline Model Description

Single-Fluid CNN (SF-CNN). A fully convolutional neural network architecture[3] with kernel size of 13 (large enough to capture the spatial span of adsorption phenomena as determined by domain

[3] Architecture details & diagram in Appendix.

Table 1: Dataset Description including experimental setting ranges of values.

Dataset Name	PW	Temp.	Pres.	MD Simulations
Single Fluid (Methane, Ethane)	[15 - 45]	[300 - 350]	[22 - 500]	1000
Two Fluid (Methane + Ethane)	[20 - 45]	[330 - 350]	[307 - 392]	100

experts). This model accepts input features in the source domain $\mathcal{D}_S$ for a particular fluid and yields the density profile at every point in the estimation horizon as specified in the inputs.

Two-Fluid CNN (TF-CNN). Similar architecture[30] to the Single-Fluid CNN model, except the outputs yield density profiles for multiple fluids (i.e., all constituent fluids) in a mixture M being modeled. Additionally the experimental settings added as input to the model also comprise the partial pressures of each constituent fluid in M.

4.3 Training & Evaluation Pipeline

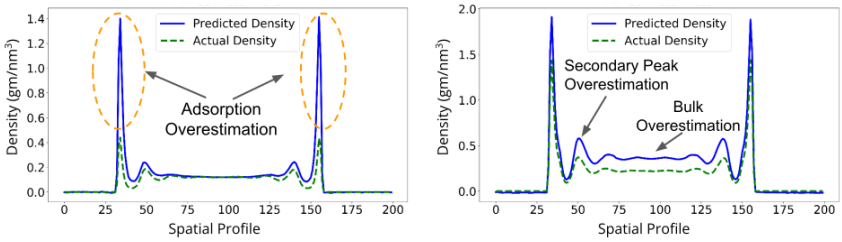
In each of our single fluid and two fluid experimentation settings we employ 64% of the data for training and 36% for testing (data sizes i.e., number of simulations in Table 1). Each model is trained for 15000 epochs with the Adam optimizer, a learning rate of $1e^{-4}$ and mean squared error loss[4].

5 Results & Discussion

We inspect our proposed NanoNet-SG framework in multiple qualitative and quantitative contexts to better understand model performance. Specifically, we are interested in dissecting the following aspects of model performance.

(1) The need for Domain Adaptation. (2) Overall generalization performance. (3) Effect of *science guidance* on domain consistency.

5.1 The Need for Domain Adaptation



(a) SF-CNN (Methane) (b) SF-CNN (Ethane)

Figure 2: Sample single fluid prediction characterization (SF-CNN) of fluid mixture without inductive transfer learning. SF-CNN models yield overestimated predictions (blue curve) of mixture densities both for methane and ethane.

To investigate this aspect, we leverage Single-Fluid CNN models (i.e., trained on single fluid simulations) to directly yield predictions in the target domain (i.e., without re-training). Fig. 2 depicts a qualitative sample[5] (pore width = 20) of the prediction performance of the Single-Fluid CNN models (trained on large corpus of single fluid simulations in similar experimental settings, separately for methane & ethane) tested on fluid density prediction in the target domain.

We can notice distinct deviations in prediction performance in the Single-Fluid CNN model trained on individual fluids. In the case of the methane prediction curve, we notice that the Single-Fluid CNN model is able to predict the bulk behavior of the fluid accurately while overestimating the adsorption effects at the pore boundaries. However, in the case of ethane prediction, we notice that the bulk density behavior of the fluid is overestimated by the corresponding Single-Fluid CNN model

[4]NanoNet-SG of course also employs the science-guided loss as in Eq. 1.

[5]More samples in appendix.

while the overestimation of the primary peaks (i.e., the highest peak at each end of the pore boundary) is relatively minimal. Thus, Fig. 2 reiterates the need for adapting the predictions of the Single-Fluid CNN model to also account for non-trivial *mixture effects* on each constituent fluid.

5.2 Overall generalization performance

In an effort to effectively model fluid mixtures by accounting for *mixture effects* under data paucity, we have developed the NanoNet-SG framework. Here, we present generalization results of the proposed framework compared to the Single-Fluid CNN and Two-Fluid CNN models (detailed in sec. 4.2). In Table 2, we notice that NanoNet (variant of NanoNet-SG without science guided loss)

Table 2: Generalization performance overall, across all temperatures, pressures (Pres.) and pore width (PW) ranges (i.e., as indicated in Table 1) and also in the contexts of unseen pore widths and pressures. In all cases, we notice that models that employ inductive knowledge transfer (i.e., NanoNet and NanoNet-SG) outperform the models without knowledge transfer and that NanoNet-SG with science-guidance is best performing overall.

Model Name	**Generalization Performance** (MAPE)		
	Overall (std.)	Unseen PW (std.)	Unseen Pres. (std.)
SF-CNN	42.52%(0.77)	40.66%(3.14)	43.69%(1.56)
TF-CNN	10.00%(0.54)	11.58%(3.75)	11.05%(1.50)
NanoNet	5.22%(1.00)	5.04%(**1.04**)	5.40%(0.769)
NanoNet-SG	**4.49**%(**0.25**)	**4.56**%(1.10)	**4.65**%(**0.41**)

and NanoNet-SG outperform other baselines thereby exhibiting superior generalization performance. Specifically we notice that the models exhibit better generalization performance overall as well as in the context of unseen pore widths and pressures. We also notice that NanoNet & NanoNet-SG models yield more consistent predictions (i.e., low standard deviations) across all inspected generalization contexts.

From Table 2 , we can infer that inductive transfer has a significant effect in domain adaptation and enables improved generalization performance . We further infer that NanoNet-SG model is the best performing overall (also with the least overall standard deviation - Table. 2) primarily due to science-guidance incorporated during model training leading to improved prediction consistency. Further baseline comparison (with state of the art regression baselines) are included in the appendix.

Qualitative Performance Characterization

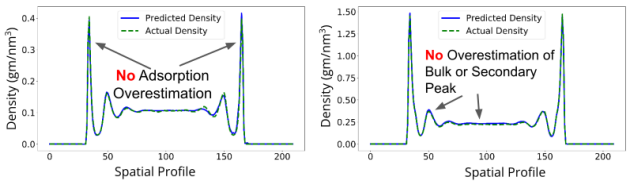


Figure 3: Sample of predictions by NanoNet-SG (blue curve) overlaid with corresponding ground-truth density curves (dotted green line) produced by a molecular dynamics (MD) simulation. Fig. 3a shows results for methane for a small pore (i.e., 21 Å) and Fig. 3b showcases similar results for ethane. We notice significant improvements in bulk and peak overestimation compared to SF-CNN (Fig. 2).

Fig. 3 demonstrates (qualitatively)[6] the improvements in fluid density estimation obtained for each of the two fluids (i.e., methane and ethane). Specifically, we notice from Fig. 3a, Fig. 3b, a significant improvement in the prediction of the peak densities and bulk densities. This is in contrast to the prediction results by the SF-CNN models without science guidance or ITL wherein peak and bulk

[6]More qualitative results in appendix.

densities were significantly overestimated. Hence, based on results in Fig. 3 and result in Table 2we may infer the effectiveness of our proposed NanoNet-SG model.

5.3 Effect of Science Guidance on Domain Consistency

In addition to prediction performance characterization, we also investigate domain consistency of NanoNet-SG . In Fig. 4 we plot the evolution of cumulative[7] total density over pressure (across the entire range of interest of pressure) at different pore widths and temperatures. Each subplot Fig. 4(a) - (c) compares the total cumulative density derived from the density prediction of NanoNet (yellow curve) and NanoNet-SG (blue curve). The phenomenon of interest is to ensure that the evolution of total cumulative density depicts monotonically non-decreasing behavior according to known scientific theory (Remark 1). In each plot in Fig. 4 we notice that the evolution of the total cumulative density derived from the NanoNet predictions violate the monotonically non-decreasing constraint at multiple points (indicated with dotted red circles) while the corresponding regions exhibit consistency with the domain knowledge for the NanoNet-SG cumulative densities (dotted green circles).

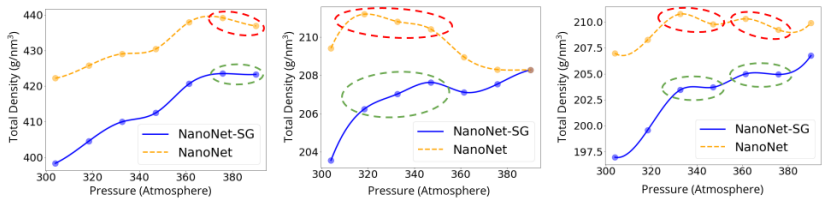


Figure 4: Each figure indicates total cumulative density evolution across a range of pressure values for specific pore width, temperature combinations. Yellow curve shows cumulative total density derived from NanoNet predictions and blue curve indicates cumulative total density derived from NanoNet-SG prediction. Red, green dotted circles highlight violations and consistency respectively of monotonically non-decreasing thermodynamic constraint detailed in Remark 1.

6 Conclusion

In this work, we have presented NanoNet-SG , a novel model for science-guided learning under data paucity via. inductive transfer learning. Our contributions are the novel thermodynamic constraint encoded into NanoNet-SG and the inductive transfer learning framework for modeling density of fluid mixtures under nano-confinement and data paucity. Through rigorous analysis, we have demonstrated that NanoNet-SG yields significantly better predictions than baselines for fluid density estimation. The superior generalization performance of NanoNet-SG holds in both the *bulk* and *peak* regions[8] of the density profile and also in the case of seen and unseen settings (i.e., unseen pore widths and pressures). Finally, we also demonstrate that NanoNet-SG yields more domain-consistent predictions as a function of the novel science-guided training mechanism.

References

Shohreh Amini and Shahab Mohaghegh. Application of machine learning and artificial intelligence in proxy modeling for fluid flow in porous media. *Fluids*, 4(3):126, 2019.

Steven Brunton, Bernd Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *arXiv preprint arXiv:1905.11075*, 2019.

Sihong Chen, Kai Ma, and Yefeng Zheng. Med3d: Transfer learning for 3d medical image analysis. *arXiv preprint arXiv:1904.00625*, 2019.

Priyanka Gupta, Pankaj Malhotra, Jyoti Narwariya, Lovekesh Vig, and Gautam Shroff. Transfer learning for clinical time series analysis using deep neural networks. *Journal of Healthcare Informatics Research*, 4(2):112–137, 2020.

[7] summation of total density of each constituent fluid in the mixture

[8] see appendix

Wenhui Li, Yiling Nan, Zilin Zhang, Qing You, and Zhehui Jin. Hydrophilicity/hydrophobicity driven co2 solubility in kaolinite nanopores in relation to carbon sequestration. *Chemical Engineering Journal*, 398:125449, 2020a.

Xuhong Li, Yves Grandvalet, Franck Davoine, Jingchun Cheng, Yin Cui, Hang Zhang, Serge Belongie, Yi-Hsuan Tsai, and Ming-Hsuan Yang. Transfer learning in computer vision tasks: Remember where you come from. *Image and Vision Computing*, 93:103853, 2020b.

Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

Nicholas Lubbers, Animesh Agarwal, Yu Chen, Soyoun Son, Mohamed Mehana, Qinjun Kang, Satish Karra, Christoph Junghans, Timothy C Germann, and Hari S Viswanathan. Modeling and scale-bridging using machine learning: Nanoconfinement effects in porous media. *Scientific Reports*, 10(1):1–13, 2020.

Paulo JM Monteiro, Chris H Rycroft, and Grigory Isaakovich Barenblatt. A mathematical model of fluid and gas flow in nanoporous media. *Proceedings of the National Academy of Sciences*, 109 (50):20309–20313, 2012.

Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Sebastian Ruder, Matthew E Peters, Swabha Swayamdipta, and Thomas Wolf. Transfer learning in natural language processing. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials*, pages 15–18, 2019.

Javier E Santos, Mohammed Mehana, Hao Wu, Masa Prodanovic, Qinjun Kang, Nicholas Lubbers, Hari Viswanathan, and Michael J Pyrcz. Modeling nanoconfinement effects using active learning. *The Journal of Physical Chemistry C*, 124(40):22200–22211, 2020.

Rui Wang, Karthik Kashinath, Mustafa Mustafa, Adrian Albert, and Rose Yu. Towards physics-informed deep learning for turbulent flow prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1457–1466, 2020.

Appendix
Knowledge-Guided Transfer Learning for Modeling Subsurface Phenomena Under Data Paucity

Nikhil Muralidhar
Department of Computer Science
Stevens Institute of Technology
Hoboken, NJ 07030
nmurali1@stevens.edu

Nicholas Lubbers
Los Alamos National Labs
Los Alamos, NM 87545
nlubbers@lanl.gov

Mohamed Mehana
Los Alamos National Labs
Los Alamos, NM 87545
mzm@lanl.gov

Naren Ramakrishnan
Department of Computer Science
Virginia Tech Arlington, VA 22203
naren@cs.vt.edu

Anuj Karpatne
Department of Computer Science
Virginia Tech Blacksburg, VA 24061
karpatne@vt.edu

1 Baseline Model Architecture Description

We employ the same basic architecture w.r.t tasks (i.e., density profile shape prediction, density profile bulk prediction) for our Single-Fluid CNN and Two-Fluid CNN models as detailed in Fig. 1. It must be noted that NanoNet and NanoNet-SG models employ predictions from the Single-Fluid CNN models as pre-trained baselines in their pipeline.

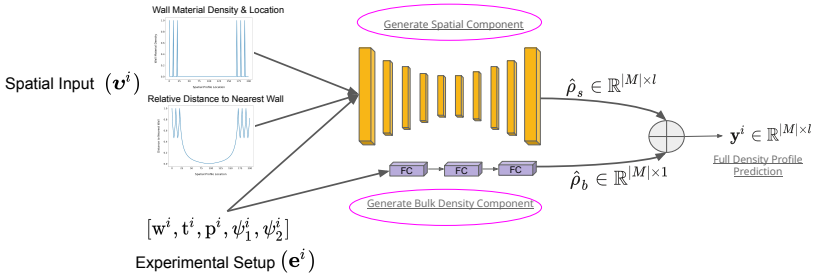


Figure 1: Architecture diagram of generic Single-Fluid CNN , Two-Fluid CNN models.

Spatial Density Profile Component. In Fig. 1, we notice that our model comprises a multi-layer fully convolutional component (depicted as yellow rectangles) which accepts the spatial pore encoding $\boldsymbol{\upsilon}^i$(see section 3.1 for details) as well as the experimental setting $\mathbf{e}^i$ (see section 3.1) as input and yields $\hat{\rho}_s \in \mathbb{R}^{|M|\times l}$ the spatial component of the density profile estimate for each fluid in the mixture M and each point in the estimation horizon l. We employ a constant kernel size of *13* and a constant number of channels *32* for both the Single-Fluid CNN and Two-Fluid CNN models.

Bulk Density Profile Component. Further, the second component of the model (Fig. 1) is a multi-layer fully connected network (depicted as purple boxes) which is conditioned only upon the experimental setup $\mathbf{e}^i$ to yield an estimate of $\hat{\rho}_b \in \mathbb{R}^{|M|\times 1}$ the *bulk* component of the density profile estimate for each fluid in the mixture M.

NeurIPS 2022 AI for Science Workshop.

The outputs $\hat{\rho}_s$ and $\hat{\rho}_b$ from the spatial and bulk components are summed together (point-wise) to yield the final density profile estimate $y^i \in \mathbb{R}^{|M| \times l}$.

Note, in the case of Single-Fluid CNN (i.e., single fluid predictions) $|M| = 1$ and the experimental setup does not comprise of partial pressures (i.e., $\mathbf{e}^i = [\mathrm{w}^i, \mathrm{t}^i, \mathrm{p}^i]$). Both Single-Fluid CNN and Two-Fluid CNN models have about $\sim$120,000 learnable parameters. The Single-Fluid CNN and Two-Fluid CNN model architectures have been developed as part of a parallel work by our group and the publication (in preparation) will be cited in the current paper upon acceptance.

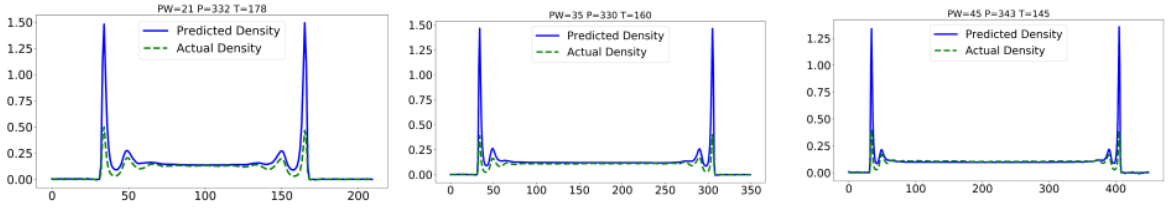
1.1 NanoNet-SG Model Architecture Details

The *Spatial Encoding Module* (sec. 3.3) and *Interaction Encoding Module* (sec. 3.3) both comprise of 2 input channels and 16 output channels. The outputs of the *Spatial Encoding* and *Interaction Encoding* modules are concatenated (channel-wise) and the 32 channel input is operated on by the *Adaptation Module* (sec. 3.3) comprising two convolutional blocks sequentially connected. First of the two convolutional blocks has 32 input channels and 16 output channels while the second convolutional block has 16 input channels followed by $|M|$ (i.e., the number of fluids in the mixture being predicted) output channels[1]. The $|M|$ channel output of the *Adaptation Module* is corrected by the output of the *Mapper* module (which is a $|e^i| \times l * |M|$ dimensional fully connected network) to produce an $|M|$ channel output to correct each point in the prediction horizon 'l' obtained from the *Adaptation Module*. Through this process, NanoNet-SG architecture is able to adapt the predictions obtained from pre-trained single fluid models to the complex fluid-mixture modeling context with only 5108 learnable parameters while also yielding improved domain consistency and superior generalization ability relative to models that are agnostic to science guidance.

2 Qualitative Figures

2.1 Single-Fluid CNN Model on Two-Fluid Prediction Task

We notice from Fig. 2, Fig. 3, that the behavior of the Single-Fluid CNN model is consistent with that presented in the main paper for other sample prediction scenarios in the target domain. Essentially, we once again notice an overestimation of the bulk density in the context of ethane profile prediction and overestimation of the peak in the methane profile prediction context.



(a) PW= 21 (Ang.), T = 332 (K), Pres. = 178 (Atm)

(b) PW= 35 (Ang.), T = 330 (K), Pres. = 160 (Atm)

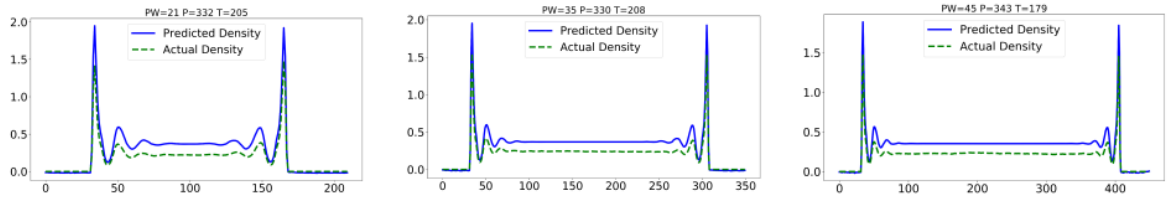
(c) PW= 45 (Ang.), T = 343 (K), Pres. = 145 (Atm)

Figure 2: Single-Fluid CNN Methane Density Profile Prediction

2.2 NanoNet-SG Model Predictions

We also include qualitative samples of the NanoNet-SG density profile predictions. Fig. 4 and Fig. 5 demonstrate that NanoNet-SG is able to yield consistent accurate predictions across all the predictions contexts sans any underestimation or overestimation behavior as observed in the case of the Single-Fluid CNN models.

[1] In our experiments $|M| = 2$

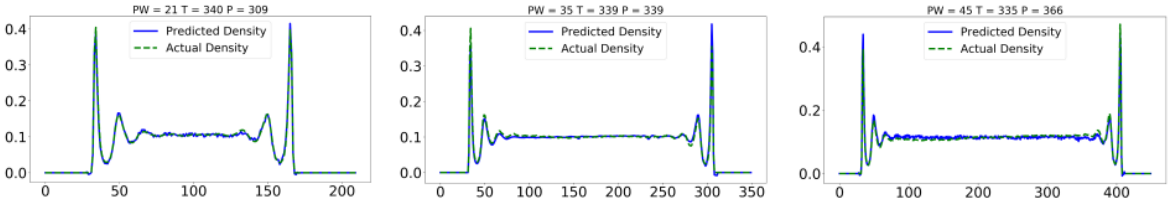


(a) PW= 21 (Ang.), T = 332 (K), Pres. = 205 (Atm)

(b) PW= 35 (Ang.), T = 330 (K), Pres. = 208 (Atm)

(c) PW= 45 (Ang.), T = 343 (K), Pres. = 179 (Atm)

Figure 3: Single-Fluid CNN Ethane Density Profile Prediction

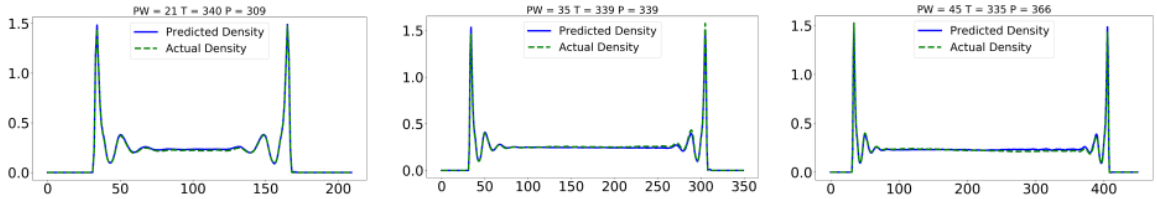


(a) PW= 21 (Ang.), T = 340 (K), Pres. = 309 (Atm)

(b) PW= 35 (Ang.), T = 339 (K), Pres. = 339 (Atm)

(c) PW= 45 (Ang.), T = 335 (K), Pres. = 366 (Atm)

Figure 4: NanoNet-SG Methane Density Profile Prediction



(a) PW= 21 (Ang.), T = 340 (K), Pres. = 309 (Atm)

(b) PW= 35 (Ang.), T = 339 (K), Pres. = 339 (Atm)

(c) PW= 45 (Ang.), T = 335 (K), Pres. = 366 (Atm)

Figure 5: NanoNet-SG Ethane Density Profile Prediction

3 Baseline Performance Comparison

3.1 Baseline Model Description

Random Forest (RF-Reg.). We train a random forest regressor (with 100 estimators) for the task of density profile prediction. Specifically, for a particular experimental setup $\mathbf{e}^i$, and a particular pore location 'q', the model is trained to predict $\hat{\mathbf{y}}_T^i(q) \in \mathbb{R}^{1\times 1}$ (i.e., the fluid density at location 'q'). The model is conditioned upon $\mathbf{e}^i$ and $\boldsymbol{\upsilon}^i(q)$ where $\boldsymbol{\upsilon}^i(q)$ represents the spatial inputs at pore location 'q'.

Gradient Boosting (GB-Reg.). The problem setup for the GB-Reg. model (with 100 estimators, squared error loss function, learning rate of 0.1) is identical to that of the RF-Reg. model (i.e., predicting density of a fluid at each location 'q' conditioned upon $\mathbf{e}^i$, $\boldsymbol{\upsilon}^i(q)$).

Linear Regression (Lin-Reg.). The problem setup for the Lin-Reg. model is identical to the RF-Reg., GB-Reg. setups.

Single-Fluid CNN (SF-CNN). A fully convolutional neural network architecture[2] with kernel size of 13 (large enough to capture the spatial span of adsorption phenomena as determined by domain experts). This model accepts input features in the source domain $\mathcal{D}_S$ for a particular fluid and yields the density profile at every point in the estimation horizon as specified in the inputs.

[2] Architecture details & diagram in Appendix.

Two-Fluid CNN (TF-CNN). Similar architecture[20] to the Single-Fluid CNN model, except the outputs yield density profiles for multiple fluids (i.e., all constituent fluids) in a mixture M being modeled. Additionally the experimental settings added as input to the model also comprise the partial pressures of each constituent fluid in M.

Performance Comparison. To further investigate the performance of NanoNet-SG, we compare

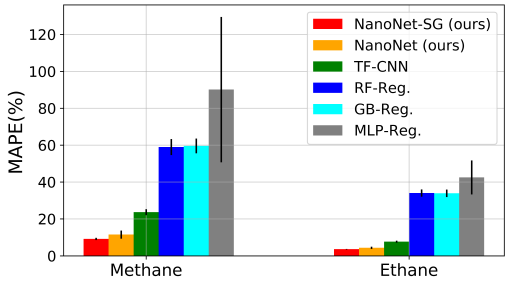


Figure 6: Prediction performance comparison with state-of-the-art regression models. NanoNet (orange), NanoNet-SG (red) models yield lowest errors and standard deviations.

the prediction performance to state-of-the-art regression baselines (details in sec. 3.1). In Fig. 6 we notice that NanoNet and NanoNet-SG models yield significantly improved prediction performance on both the methane and ethane profile prediction tasks in the fluid mixture density profile prediction context. We specifically report methane and ethane profile prediction errors for each model in Fig. 6 and notice that NanoNet and NanoNet-SG outperform all other models in both prediction tasks (i.e., methane, ethane density profile prediction).

Model Name	**Generalization Performance Breakdown** (MAPE)			
	Methane Bulk	Ethane Bulk	Methane Profile	Ethane Profile
SF-CNN	6.29%	56.52%	50.50%	56.68%
TF-CNN	4.90%	3.73%	23.69%	7.69%
NanoNet	**2.53%**	2.45%	11.56%	4.36%
NanoNet-SG	2.98%	**2.15%**	**9.19%**	**3.65%**

Table 1: Generalization performance per fluid over the bulk prediction and profile prediction contexts. We once again notice that NanoNet-SG yields best overall performance due to knowledge transfer coupled with science-guidance.

In Table 1, we further showcase a task breakdown of the generalization performance observed in Table 1 (main paper) by inspecting generalization errors in the bulk and profile prediction context per fluid for each model. We can infer from Table 1 that models NanoNet and NanoNet-SG yield significant performance improvement owing to the effect of the knowledge transfer architecture and the science guidance in the case of NanoNet-SG. Specifically, we notice that science guidance results in improved overall generalization as well as bulk and profile generalization in both the case of Methane and Ethane. Science guidance leads to NanoNet-SG gaining a **16.26%** performance improvement over its variant without science guidance (i.e., NanoNet) and significantly larger gains in performance over SF-CNN and TF-CNN models.

Qualitative Performance Characterization. We notice from Fig. 7 that NanoNet-SG yields highly accurate density estimates across the entire range of the prediction spectrum of fluid density for each of the fluids in the mixture. Also apparent from the plots is the ability of the model to fit the multiple distributional modes in the density profile owing to the density in the *peak* and *bulk* regions due to the presence and absence of adsorption effects respectively. These multiple modes are indicated by disjoint clusters of points in each plot in Fig. 7. We also report the R^2 statistic for density prediction of NanoNet-SG which indicates consistent performance across the two fluids with $\mathbf{R^2}$**=0.991** for *methane* density prediction and $\mathbf{R^2}$**=0.998** for *ethane* density prediction.

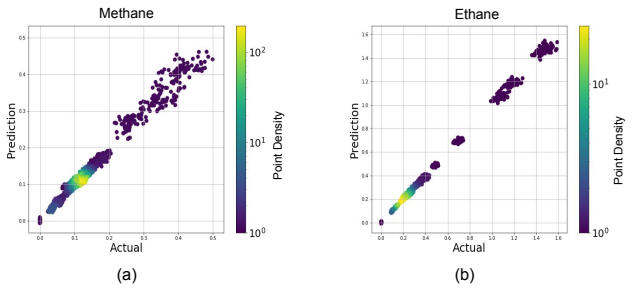


Figure 7: Qualitative performance characterization of NanoNet-SG in density profile prediction for Methane-Ethane fluid mixture. Color bar indicates distribution of points across prediction range.

3.2 Effect of Science Guidance on Domain Consistency

Further, in Fig. 8 (a), 8(b) we depict evolution of total cumulative densities for NanoNet and NanoNet-SG at a sample pore width of 25 at various temperatures. We once again notice that the evolution for NanoNet showcases significant deviations from the monotonically non-decreasing domain constraint whereas the NanoNet-SG model predictions result in total density curves that adhere to the scientific domain knowledge much more consistently. From Fig. **??** and Fig. 8, we can infer that the NanoNet-

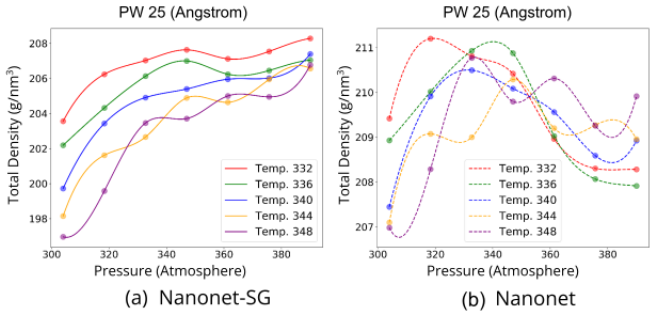


Figure 8: Each of Fig. 8(a) and 8(b) depict total cumulative density evolution with pressure (at multiple temperature values) for NanoNet-SG and NanoNet respectively at pore width 25Å.

SG model yields significantly more domain consistent results owing to the science-guided training criterion.

4 Related Work

Transfer Learning -Transfer learning as a model training paradigm has been effectively leveraged in applications like natural language processing Ruder et al. [2019], computer vision Li et al. [2020], healthcare Gupta et al. [2020], Chen et al. [2019] and many other contexts. However, few efforts have employed transfer learning under data paucity in scientific applications. To the best of our knowledge, our NanoNet-SG is the first science-guided model to employ inductive transfer learning to counter the effects of data paucity in scientific applications.

Fluid-Flow Modeling. Various ML models have been proposed for modeling fluid flows Brunton et al. [2019], Wang et al. [2020], Raissi et al. [2019]. However all these models (unlike NanoNet-SG) either assume complete knowledge of the physics (e.g., PDE, ODE, initial, boundary conditions) governing the system or employ large datasets in target domain.

Modeling Under Nano-confinement. The works Santos et al. [2020], Lubbers et al. [2020] are closest to ours. In Santos et al. [2020], authors propose an active-learning framework for training a ML emulator of MD simulations and Lubbers et al. [2020] develop an ML model to bridge fine-grained (MD) and coarse-grained (Lattice-Boltzman) simulations of nano-confined fluids. However, neither of these works tackle the complex problem of modeling fluid mixtures and neither work employs explicit science guidance in the loss function for modeling under data paucity. How is our

work different? 1. Uses weak scientific domain knowledge. 2. Modeling under data paucity 3. No other works leverage inductive transfer learning and science guidance. 4. Our work is different from Javier's work as we do not assume access to the MD simulator (only to the scientific domain knowledge and initial data).

References

Steven Brunton, Bernd Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *arXiv preprint arXiv:1905.11075*, 2019.

Sihong Chen, Kai Ma, and Yefeng Zheng. Med3d: Transfer learning for 3d medical image analysis. *arXiv preprint arXiv:1904.00625*, 2019.

Priyanka Gupta, Pankaj Malhotra, Jyoti Narwariya, Lovekesh Vig, and Gautam Shroff. Transfer learning for clinical time series analysis using deep neural networks. *Journal of Healthcare Informatics Research*, 4(2):112–137, 2020.

Xuhong Li, Yves Grandvalet, Franck Davoine, Jingchun Cheng, Yin Cui, Hang Zhang, Serge Belongie, Yi-Hsuan Tsai, and Ming-Hsuan Yang. Transfer learning in computer vision tasks: Remember where you come from. *Image and Vision Computing*, 93:103853, 2020.

Nicholas Lubbers, Animesh Agarwal, Yu Chen, Soyoun Son, Mohamed Mehana, Qinjun Kang, Satish Karra, Christoph Junghans, Timothy C Germann, and Hari S Viswanathan. Modeling and scale-bridging using machine learning: Nanoconfinement effects in porous media. *Scientific Reports*, 10(1):1–13, 2020.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Sebastian Ruder, Matthew E Peters, Swabha Swayamdipta, and Thomas Wolf. Transfer learning in natural language processing. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials*, pages 15–18, 2019.

Javier E Santos, Mohammed Mehana, Hao Wu, Masa Prodanovic, Qinjun Kang, Nicholas Lubbers, Hari Viswanathan, and Michael J Pyrcz. Modeling nanoconfinement effects using active learning. *The Journal of Physical Chemistry C*, 124(40):22200–22211, 2020.

Rui Wang, Karthik Kashinath, Mustafa Mustafa, Adrian Albert, and Rose Yu. Towards physics-informed deep learning for turbulent flow prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1457–1466, 2020.