Uncertainty Quantification in Polysilicon-based MEMS: a Representation Learning Comparison

José Pablo Quesada-Molina [1,2] , Stefano Mariani [1] (Ph.D. Supervisor)
josepablo.quesada@polimi.it stefano.mariani@polimi.it
[1] Department of Civil and Environmental Engineering, Politecnico di Milano (Italy)
[2] Department of Mechanical Engineering, Universidad de Costa Rica (Costa Rica)

MASCOT-NUM 2022
Clermont Ferrand, France
7-9 June 2022

POLITECNICO MILANO 1863

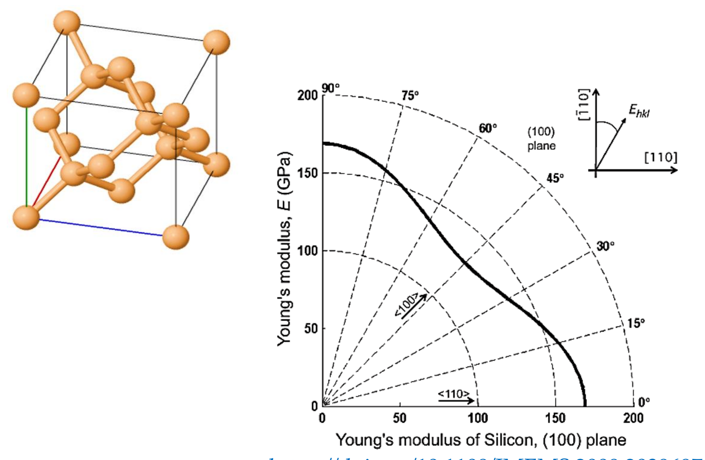
UNIVERSIDAD DE COSTA RICA

Objectives of the work

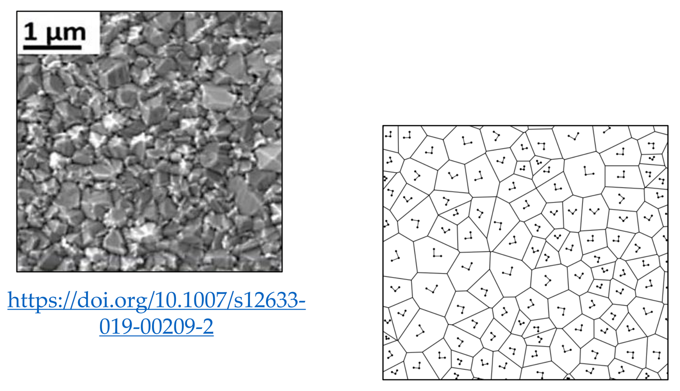
- Implement a deep learning-based UQ framework capable to account for material-related uncertainties affecting the reliability of polysilicon MEMS in a straightforward data-driven manner.
- Explore the potential to enhance the performance of the implemented deep learning model via a transfer learning strategy, exploiting a different representation learning approach for the feature extraction.

Background

- Monocrystalline silicon crystallizes in a diamond cubic crystal structure which confers anisotropic elasticity properties.
- The intrinsic heterogeneity of polycrystalline silicon induce scattering of the effective mechanical properties in the structural components of MEMS.
- Due to the stochastic nature of the variables controlling the final microstructure, statistical analyses are required to characterize the variability ranges in the effective properties for the reliable design of these microscopic devices.



https://doi.org/10.1109/JMEMS.2009.2039697



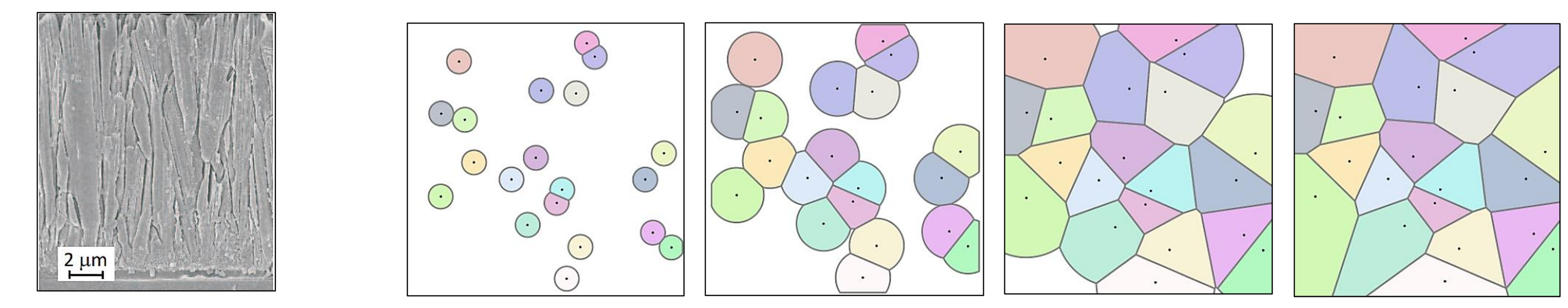
https://doi.org/10.1007/s12633-019-00209-2

https://doi.org/10.1615/intjmultcompeng.v9.i3.50

Methodology

Input data generation

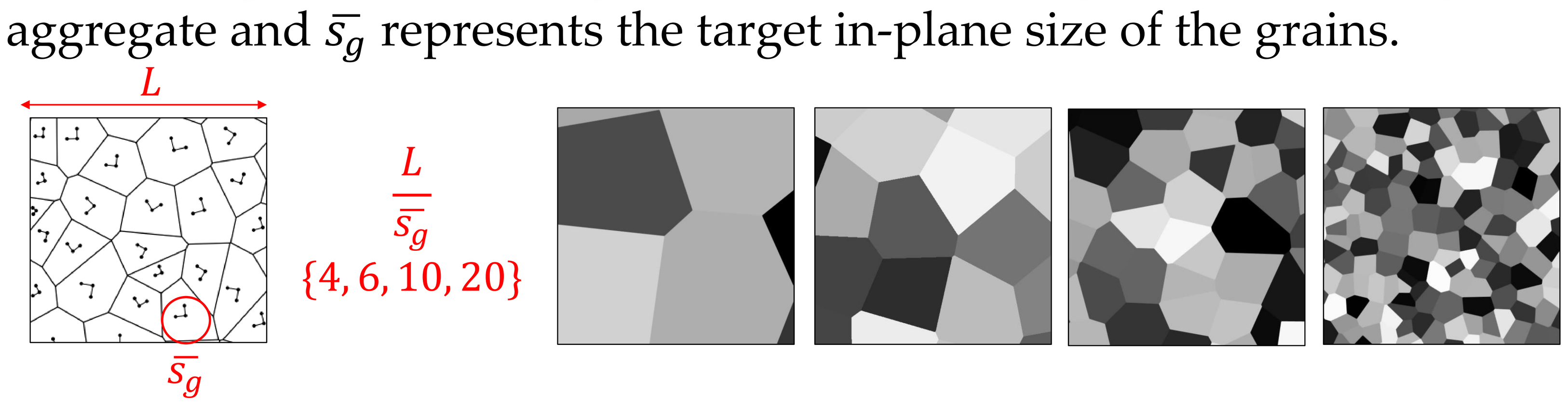
- Following the approach presented in former works[1] , **stochastic volume elements (SVE)** representing cross-sections of epitaxially grown polysilicon thin-films have been generated.



https://doi.org/10.1615/intjmultcompeng.v9.i3.50

https://community.wolfram.com/groups/-/m/t/1202074

- Different datasets have been assembled **featuring small ratios $L/\bar{s}_g$**, where L represents the length-scale characterizing the size of the grain aggregate and $\bar{s}_g$ represents the target in-plane size of the grains.

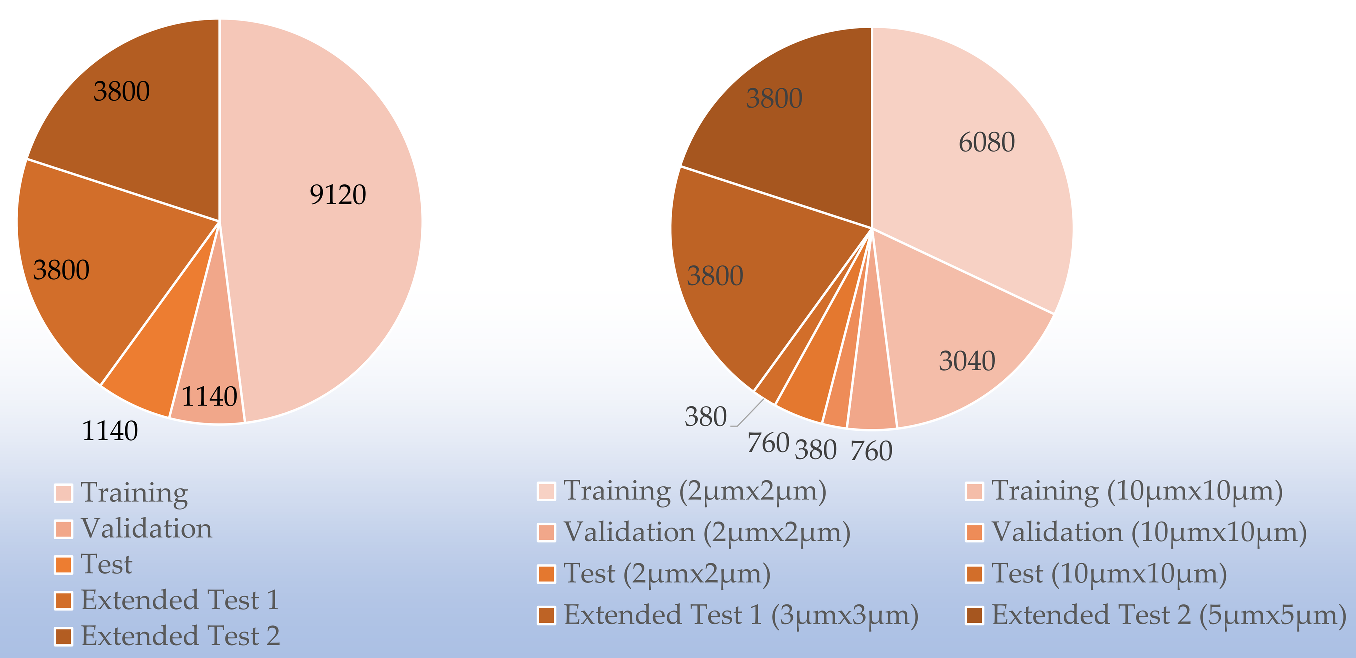


- Generated data has been formatted as single-channel images with a resolution of 128 px × 128 px, wherein, the grey level represents the in-plane lattice orientation of each grain.
- Effective elastic properties of epitaxially grown polysilicon thin-films e.g., the Young's modulus, E are described by $LogN(\mu, \sigma)$ characterized by scale-dependent μ, σ values.
- Scale-dependent statistical indicators describing the target values of the generated data are shown in the table:

SVE size	μ [GPa]	σ [GPa]
2μm × 2μm	149.8	5.4
3μm × 3μm	149.6	3.9
5μm × 5μm	149.4	2.4
10μm × 10μm	149.0	1.3

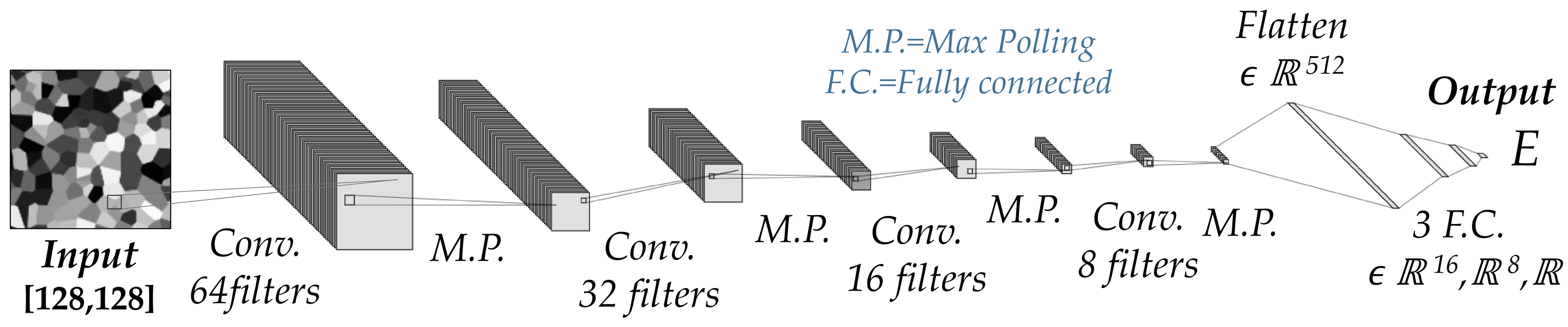
Datasets composition

- Unbalanced datasets aim to better capture the larger variation of the target values associated with the smaller characteristic SVE sizes.



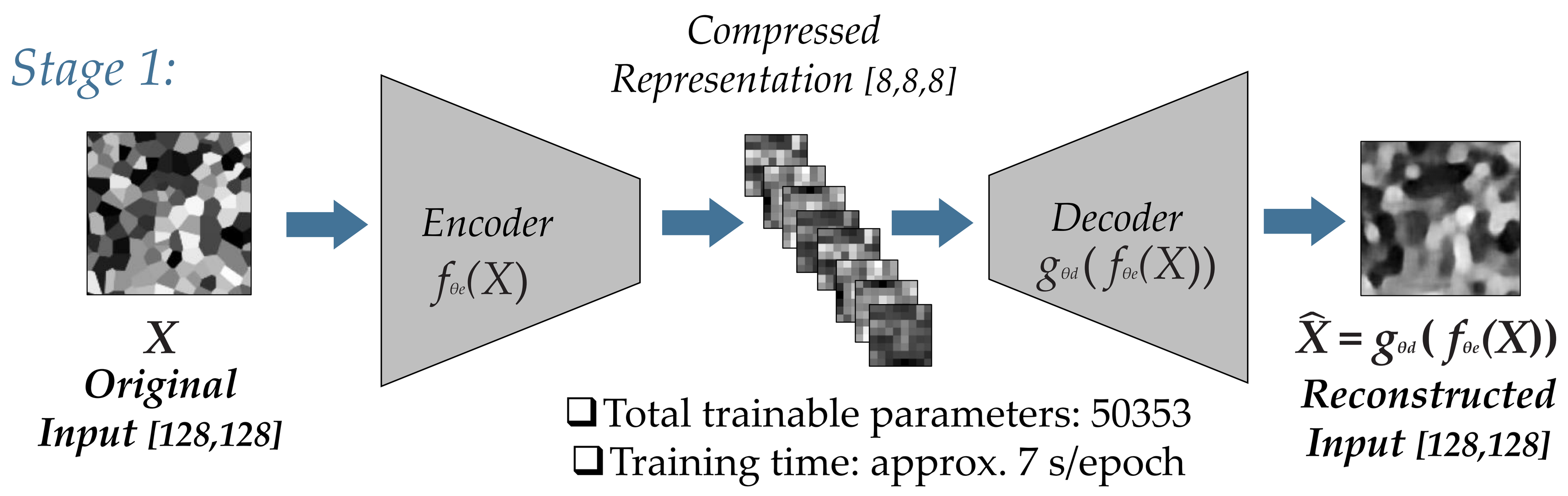
Representation learning comparison

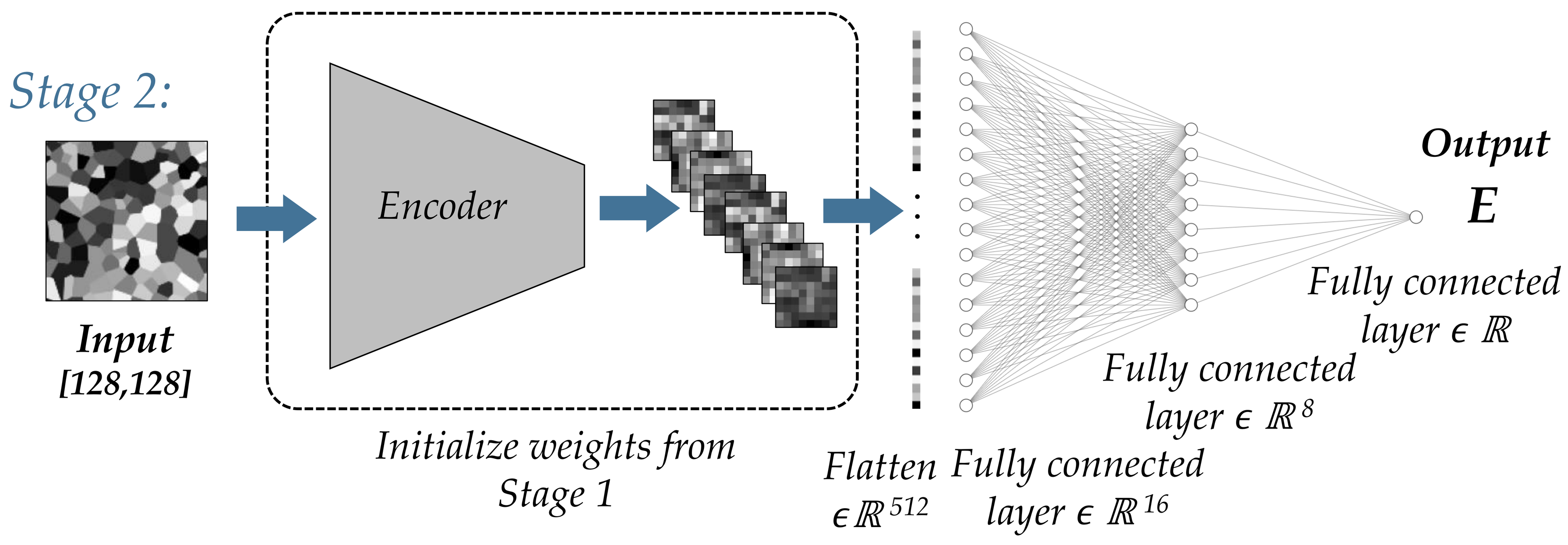
- A **convolutional neural network (CNN)** is first optimized as a baseline to provide the mapping between the images of polysilicon microstructure and homogenized Young's modulus E.



- A convolutional **autoencoder (AE) neural network** is pre-trained and employed for transfer learning purposes:

$$J(\theta, X) = L(X, g_{\theta d}(f_{\theta e}(X)))$$



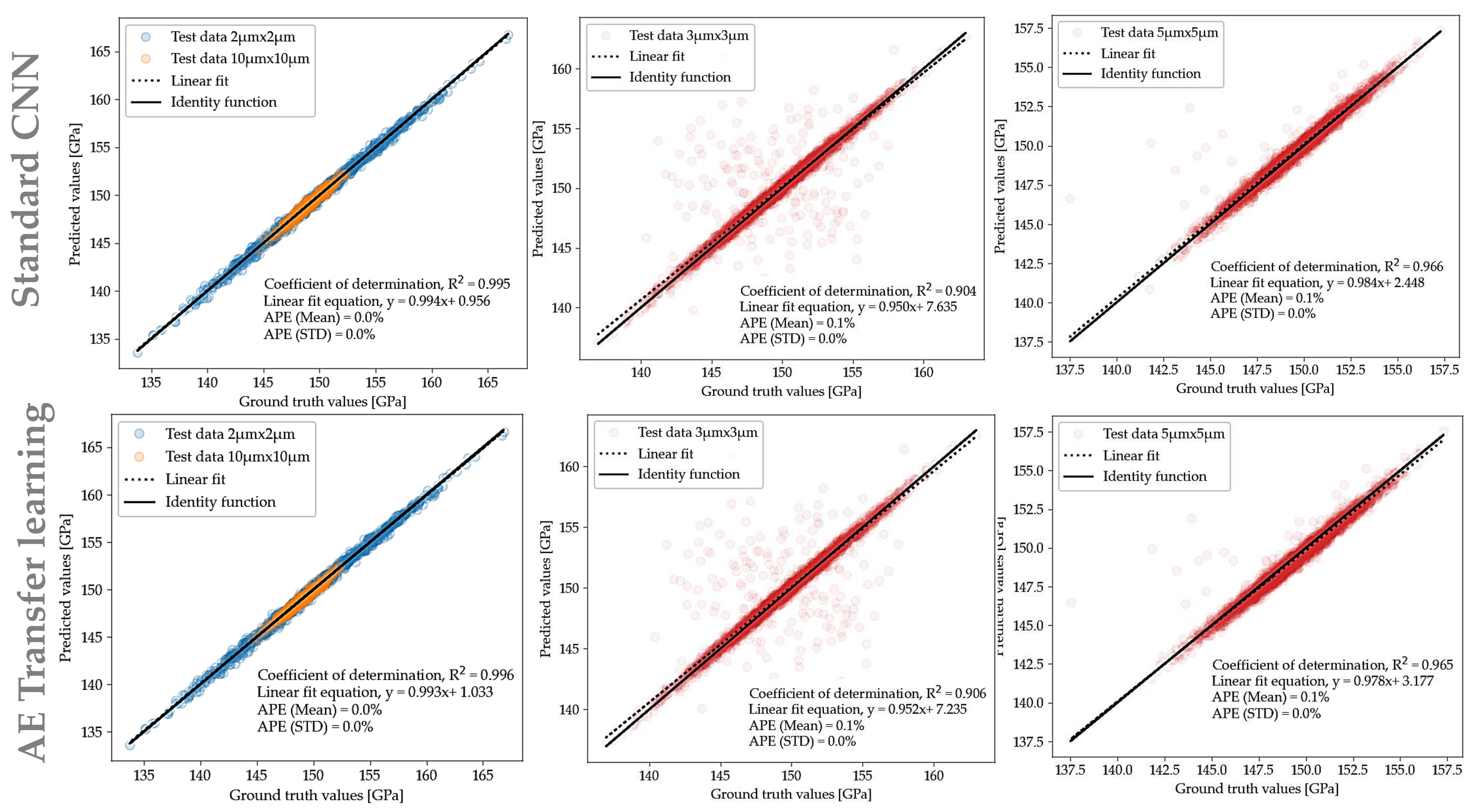


(I) Fix the weights of the encoder after Stage 1 and update the weights of the fully connected layers only.
(II) Allow the update of the entire network (encoder + fully connected layers).

General implementation details:
- ❑Optimizer: Adamax ($\eta = 5\text{x}10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 1\text{x}10^{-7}$)
- ❑ReduceLROnPlateau (factor=0.2, patience =5, min $\eta = 1\text{x}10^{-4}$)
- ❑Loss: Mean Squared Error (MSE)
- ❑Convolutional kernel size: [3×3]
- ❑Pooling kernel size: [2×2]
- ❑Early stopping patience: 100 epochs
- ❑Total trainable parameters: 33241
- ❑Training time: approx. 3 s/epoch (NVIDIA GeForce RTX 3090 GPU)

Results

Predictions over test samples representative of the training data. *Predictions over test samples characterized by length-scale values different from those during training.*



Conclusions

- None of the two explored strategies have produced significantly better results (in terms of quality or convergence speed) than those obtained by the standard CNN model adopting the default weight initialization scheme.
- The effectiveness of the transfer learning strategy hinges on the degree of correlation between the successive tasks as the feature extraction process has been demonstrated to be highly target-sensitive.

References

[1] Mariani, S. et al, 2011, *Int. J. Mult. Comp. Eng*, https://doi.org/10.1615/intjmultcompeng.v9.i3.50
[2] Quesada-Molina, J.P. et al, 2020, *EuroSimE*, https://doi.org/10.1109/eurosime48426.2020.9152690
[3] Quesada-Molina, J.P. et al, 2022, *EuroSimE*, https://doi.org/10.1109/EuroSimE54907.2022.9758899