frontiers | Frontiers in Materials

TYPE Original Research
PUBLISHED 16 September 2022
DOI 10.3389/fmats.2022.971816

OPEN ACCESS

EDITED BY
Elías Cueto,
University of Zaragoza, Spain

REVIEWED BY
Alberto Badias,
Polytechnic University of Madrid, Spain
Reza Abedi,
The University of Tennessee, Knoxville, United States

*CORRESPONDENCE
Axel Aublet,
axel.aublet@minesparis.psl.eu
David Ryckelynck,
david.ryckelynck@mines-paristech.fr

SPECIALTY SECTION
This article was submitted to Computational Materials Science, a section of the journal Frontiers in Materials

RECEIVED 17 June 2022
ACCEPTED 11 August 2022
PUBLISHED 16 September 2022

CITATION
Aublet A, N'Guyen F, Proudhon H and Ryckelynck D (2022), Multimodal data augmentation for digital twining assisted by artificial intelligence in mechanics of materials.
Front. Mater. 9:971816.
doi: 10.3389/fmats.2022.971816

Multimodal data augmentation for digital twining assisted by artificial intelligence in mechanics of materials

Axel Aublet[1,2]*, Franck N'Guyen[1], Henry Proudhon[1] and David Ryckelynck[1]*

[1]MinesParis, PSL University, Centre for Material Sciences (MAT), UMR7633CNRS, Evry, France, [2]Safran Tech, Rue des Jeunes Bois, Colombes, France

Digital twins in the mechanics of materials usually involve multimodal data in the sense that an instance of a mechanical component has both experimental and simulated data. These simulations aim not only to replicate experimental observations but also to extend the data. Whether spatially, temporally, or functionally, augmentation is needed for various possible uses of the components to improve the predictions of mechanical behavior. Related multimodal data are scarce, high-dimensional and a physics-based causality relation exists between observational and simulated data. We propose a data augmentation scheme coupled with data pruning, in order to limit memory requirements for high-dimensional augmented data. This augmentation is desirable for digital twining assisted by artificial intelligence when performing nonlinear model reduction. Here, data augmentation aims at preserving similarities in terms of the validity domain of reduced digital twins. In this article, we consider a specimen subjected to a mechanical test at high temperature, where the as-manufactured geometry may impact the lifetime of the component. Hence, an instance is represented by a digital twin that includes 3D X-Ray tomography data of the specimen, the related finite element mesh, and the finite element predictions of thermo-mechanical variables at several time steps. There is, thus, for each specimen, geometrical and mechanical information. Multimodal data, which couple different representation modalities together, are hard to collect, and annotating them requires a significant effort. Thus, the analysis of multimodal data generally suffers from the problem of data scarcity. The proposed data augmentation scheme aims at training a recommending system that recognizes a category of data available in a training set that has already been fully analyzed by using high-fidelity models. Such a recommending system enables the use of a ROM-net for fast lifetime assessment *via* local reduced-order models.

KEYWORDS

high-dimensional data, model order-reduction, data pruning, oversampling, computational mechanics, lifetime prediction

1 Background

A digital twin (DT), of a system or a component, is an ultra-realistic model in geometric details Bellinger et al. (2011). These details include manufacturing anomalies. More precisely, when considering model-based engineering systems, the DT integrates ultra-high fidelity simulations, maintenance history, and all available historical among real data. This information is used to mirror the life of their respective "flying twin," as proposed in Glaessgen and Stargel (2012) for aeronautical applications. In this article, we consider real instances of mechanical components (RIMC) that undergo thermo-mechanical loadings in an experiment that reflects what an aircraft engine imposes on a high-pressure turbine blade. The as-manufactured geometries are observed by X-Ray computed tomography (CT) which generates accurate 3D digital images of RIMC. A numerical model is a DT only if its corresponding RIMC exists. Making the best use of existing data while exploiting the information with classical mathematical models is an important issue. Data-driven modeling is emerging as a powerful paradigm for physical and biological systems, according to Zhang et al. (2019). In various industrial fields related to structural mechanics, most numerical models incorporate scientific knowledge about the mechanics of materials. Hence, the numerical instance of a DT incorporates multimodal data. Each modality of data has its own ambient space related either to material parameters, geometrical and morphological representation, or boundary and initial conditions for partial differential equations. One important modality is related to the synthetic data forecast by model-based numerical simulations, predicting the lifetime of the RIMC, for instance, upon observational data. Such synthetic data support decisions in engineering for maintenance, operating optimization, or decommission of RIMC. With the development of non-destructive testing (NDT) in the manufacturing industry, we can expect a growing activity on image-based DTs (Seon et al., 2020; Launay et al., 2021a) for accurate descriptions of as-manufactured geometries and microstructural properties of structural components. Such data are crucial for lifetime predictions (Aublet et al., 2022). Unfortunately, image-based DTs in material science are so complex and time-consuming, so in practice, this task does not scale with the frequency of quality inspection in manufacturing. This is a major scaling issue for online DT of as-manufactured structural components, where lifetime prediction is strongly affected by geometrical defects. Another issue is the curse of dimensionality related to the dimension of the geometrical ambient space, so a dense sampling of this space is unaffordable. So, learning regression for lifetime prediction from observational data, as in cross-modal predictions (Launay et al., 2021a), may not be relevant for such image-based modeling. In this work, we augment multimodal data in order to train a ROM-net (Daniel et al., 2020). A ROM-net learns a piecewise linear latent space for model order-reduction. It incorporates a classifier, in its first layers, that recommends the selection of a local reduced-order model (LROM), from a digital image of the RIMC, where each LROM has its own label. In this article, we assume that the latent space has a structure of vector bundle that is locally trivial. This means that LROMs have a validity domain larger than the support of training data (Ryckelynck et al., 2015). In the sequel, DTs that have similar validity domains for their specific LROM are termed ROM-similar. Therefore, ROM-net predictions should be accurate when the input data have ROM-similar data in the dictionary of LROM. Unfortunately, the multimodal data related to image-based DT are very scarce, compared to the dimension of the ambient spaces of these data. Furthermore, we need to prove the efficiency of digital twining assisted by artificial intelligence prior to getting more data related to manufactured components. Hence ROM-net, or more precisely, the classifier of the ROM-net, needs augmented data (Lecun et al., 1998). A review on image-data augmentation schemes is available in Shorten and Khoshgoftaar (2019). Both image warping and oversampling augmentations are proposed in the literature, so labels are preserved. Here, the multimodal data involved in digital twining are required to account for the causality relation that exists between the model-based predictions and observational image data. In data warping augmentation alone, the causality relation between observational data and simulated data is not preserved. Hence, labels that refer to LROM cannot be preserved by data warping. So, we propose a dedicated oversampling procedure based on ROM-similarity. A schematic view of the assumed piecewise linear latent space is shown in **Figure 1**. In this figure, ROM-similar data are aligned on the same slope. We consider both experimental data and simulated data related to the lifetime of a realistic structural component, composed of CMSX4-PLUS super alloy, undergoing cyclic loading and plastic strains at high temperatures. The next section details the proposed scheme for multimodal data oversampling. In **Section 3**, two examples are provided. The first one is an augmentation around a defect in a 2D picture which affects the local strain fields. In the second example, we give a detailed description of the considered multimodal data for an academic example and for a mechanical component undergoing cyclic loading at high temperature for its lifetime characterization. The ambient spaces related to the multimodal data in this article are realistic regarding manufactured parts of the aeronautics industry (Aublet et al., 2022).

2 Methods

The multimodal data involved in a DT are denoted by $\boldsymbol{\mu}$ for material or metallurgical parameters, $\boldsymbol{\mu}_{\square}$ for digital images, p for the lifetime prediction, and a matrix $\mathbf{u} \in \mathbb{R}^{\mathcal{N} \times n_t}$ of model-based simulated data over n_t time steps. The physics-based

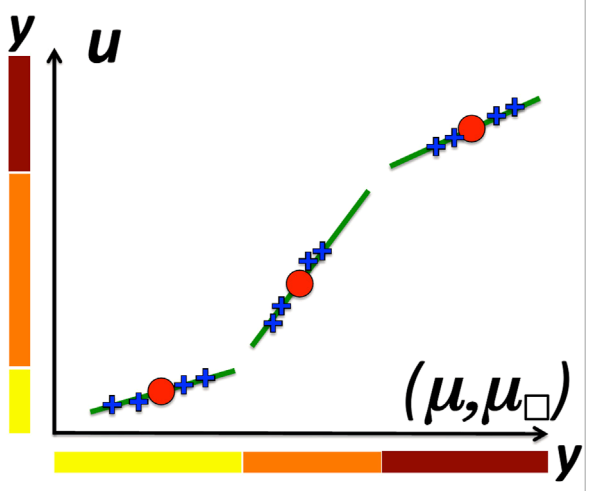

FIGURE 1
Schematic view of augmented data (blue crosses) around digital twins (red points) in a multimodal ambient space ($\boldsymbol{\mu}$, $\boldsymbol{\mu}_\square$, $\mathbf{u}$). Oversampling assumes a piecewise linear latent space (green slopes). Each linear subset of the latent space has a dedicated label y. The extent of green segments illustrates the validity domain of LROMs.

equations read as follows: find the primal variable (or state variable) $\mathbf{u}$ and the simulation outcome p for given k, $\boldsymbol{\mu}$, and $\boldsymbol{\mu}_\square$ such that,

$$\mathbf{r}\left(\mathbf{u};k,\boldsymbol{\mu},\boldsymbol{\mu}_\square\right)=0,\ k=1,\dots n_t, \qquad p-\ell\left(\mathbf{u}\right)=0 \tag{1}$$

where $\mathbf{r}(\cdot;k,\cdot,\cdot)\in\mathbb{R}^{\mathcal{N}}$, $k=1\dots n_t$ is a given series of complex physics-based equations that forecast primal variables $\mathbf{u}[:,k]$ (i.e., displacements in mechanical models) and ℓ is a much less complex function for lifetime prediction knowing $\mathbf{u}$. Residuals $\mathbf{r}$ set up the causality relation between the multimodal data. n_t is the number of residual vectors required to predict the n_t columns of $\mathbf{u}$. For instance, n_t is the number of loading steps in a mechanical problem.

The solution manifold of all $\mathbf{u}$ fulfilling Equation 1, for given $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_\square$, is denoted by E. For reduced-order modeling purposes, we assume that E has a structure that the data augmentation scheme has to preserve. This structure is a vector bundle of rank N. It is a family of vector spaces of dimension N parametrized by a base space denoted by M. In the proposed framework, M is the manifold of primal variables $\mathbf{u}$ for digital twins. The vector bundle is assumed to be locally trivial, in the sense that the local vector space is constant on subsets of M. In the sequel, these subsets of M have a dedicated label y, and this label refers to a local reduced basis $\mathbf{V}(y)\in\mathbb{R}^{\mathcal{N}\times N}$. An instance of a DT is denoted by $\mathbf{X}^{(i)}=\{\boldsymbol{\mu}^{(i)},\boldsymbol{\mu}_\square^{(i)},\mathbf{u}^{(i)},p^{(i)}\}$, with $\mathbf{u}^{(i)}\in M$. Its label is denoted by $y^{(i)}$, and $\mathbf{V}(y^{(i)})$ is the local reduced basis attached to the i^{th} DT. The reduced basis $\mathbf{V}(y^{(i)})$ contains the left eigenvectors of the truncated singular value decomposition of $\mathbf{u}^{(i)}$. It is an orthogonal reduced basis. The original set of DT indices is denoted by $\mathcal{D}_{t0}$. The set of indices related to augmented data is denoted by $\mathcal{D}_{aug}$, such that $\mathcal{D}_{t0}\cap\mathcal{D}_{aug}=\emptyset$. Augmented data $\mathbf{X}^{(a)}$, $a\in\mathcal{D}_{aug}$, are admissible, if they preserve the structure of E in the following sense: it exists $i\in\mathcal{D}_{t0}$ such that:

$$\mathbf{u}^{(a)}\in\operatorname{span}\left(\mathbf{V}\left(y^{(i)}\right)\right) \tag{2}$$

$$y^{(a)}=y^{(i)} \tag{3}$$

$$\mathbf{r}\left(\mathbf{u}^{(a)};k,\boldsymbol{\mu}^{(a)},\boldsymbol{\mu}_\square^{(a)}\right)=0,\ k=1,\dots n_t, \qquad p^{(a)}-\ell\left(\mathbf{u}^{(a)}\right)=0 \tag{4}$$

This admissibility criteria involve, in Equation 3, the usual constrains that preserve labels of original data. We acknowledge that the proposed data augmentation scheme does not improve the sampling of the base space M. Its purpose is to augment the number of instances that share the same labels. Figure 1 is illustrating the structure of manifold E. In this figure, the sampling points on manifold M are the red points. The local vector spaces are represented as green lines. The desired augmented data are represented by blue crosses. The labels have also been represented on each axis of this plot.

Usually, $\mathcal{N}$ is large and can range from 10^5 to 10^7. The higher the resolution of digital images, the higher the dimensions of $\boldsymbol{\mu}_\square$ and $\mathbf{u}$. In practice, $\mathbf{r}$ and $\mathbf{u}$ are related to a finite element mesh of the digital image $\boldsymbol{\mu}_\square$. For such high-dimensional multimodal data, preserving memory storage capabilities while performing data augmentation is the main issue to achieve feasible augmentations. The storage limits of high-dimensional multimodal data were not in the scope of previous articles on simulated data augmentation (Daniel et al., 2021; Launay et al., 2021b). Without solving this issue, no data oversampling is possible here. Therefore, to limit the memory requirement and computational time, oversampling is coupled with a data pruning technique (Hilth et al., 2019) and numerical approximations. The method intends to restrict Equations (2) and (4) to sets of selected rows among the complete set of raws. These sets are denoted, respectively, as $\overline{\mathcal{F}}$ and $\mathcal{F}$. In order to have a consistent selection of these sets of rows, the data pruning procedure accounts for the finite element description of the residual $\mathbf{r}$ and the spatial interpolation of the primal variable $\mathbf{u}$. This interpolation uses finite element shape functions denoted by $(\boldsymbol{\varphi}_j)_{j=1,\dots\mathcal{N}}$. The interpolated primal variable reads:

$$\tilde{\mathbf{u}}\left(\boldsymbol{\xi};k,\boldsymbol{\mu},\boldsymbol{\mu}_\square\right)=\sum_{j=1}^{\mathcal{N}}\boldsymbol{\varphi}_j(\boldsymbol{\xi})\ u[j,k]\left(\boldsymbol{\mu},\boldsymbol{\mu}_\square\right) \quad \forall\,\boldsymbol{\xi}\in\Omega \tag{5}$$

where Ω is the geometrical domain occupied by the DT, and $\boldsymbol{\xi}$ is the position vector in this domain. In image-based DTs, the finite element shape functions and the domain Ω are specific

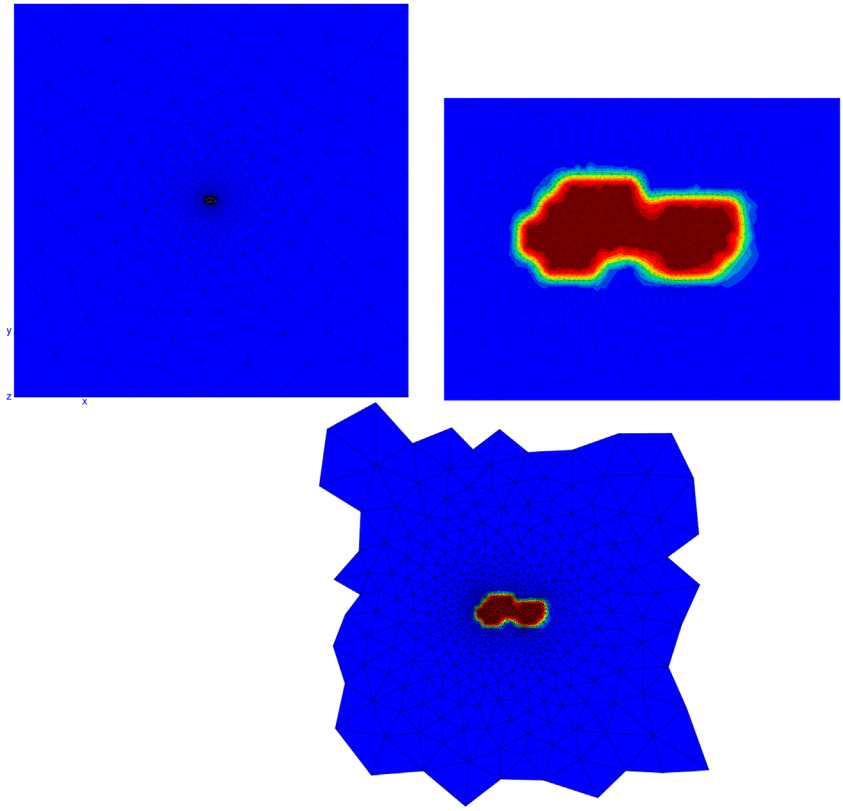

FIGURE 2
The finite element mesh of one digital twin (on the left), a focus on the phase field of the void at the center of this mesh (Right), and the mesh of the reduced domain Ω_R (Below).

to each instance of $\boldsymbol{\mu}_{\square}^{(i)}$, for i in $\mathcal{D}_{t0}$. But this dependence is not included in notations here, for the sake of simplicity. The data pruning procedure proposed in Hilth et al. (2019) introduces a reduced domain denoted by Ω_R. This reduced domain is the support of finite element shape functions related to $\mathcal{F}$:

$$\Omega_R = \cup_{j\in\mathcal{F}} \text{supp}\left(\boldsymbol{\varphi}_j\right) \tag{6}$$

and $\overline{\mathcal{F}}$ is the set of indices of the variable $\mathbf{u}$ restricted to Ω_R, so that the residual restricted to $\mathcal{F}$ can be fully computed using solely $\mathbf{u}[\overline{\mathcal{F}}]$. For the sake of simplicity, this restricted residual is denoted by $\mathbf{r}(\mathbf{u}[\overline{\mathcal{F}},:];\cdot)[\mathcal{F}]$. $\mathcal{F}$ is an user-defined pruning parameter. It is important to note that Ω_R must cover the region of interest for geometrical variations.

For pruned data, approximation errors in **Equation 2** must be tolerated. We propose a two-step augmentation technique that contains 1) a random sampling step of parameters $\boldsymbol{\mu}, \boldsymbol{\mu}_{\square}$ in the vicinity of $(\boldsymbol{\mu}^{(i)}, \boldsymbol{\mu}_{\square}^{(i)})_{i\in\mathcal{D}_{t0}}$ and 2) a screening step in accordance with the admissibility criteria for pruned data. The admissibility of pruned augmented data fulfills the following equations:

$$\boldsymbol{\gamma}^{(a)}[:,k] = \arg\min_{\mathbf{g}\in\mathbb{R}^N} \|\mathbf{u}^{(a)}\left[\overline{\mathcal{F}},k\right] - \mathbf{V}^{(i)}\left[\overline{\mathcal{F}},:\right]\mathbf{g}\|,\ k = 1,\ldots n_t \tag{7}$$

$$\eta^{(a)} = \frac{\|\mathbf{u}^{(a)}\left[\overline{\mathcal{F}},:\right] - \mathbf{V}^{(i)}\left[\overline{\mathcal{F}},:\right]\boldsymbol{\gamma}^{(a)}\|_F}{\|\mathbf{u}\left[\overline{\mathcal{F}},:\right]\|_F} \tag{8}$$

$$s = 1 - \eta^{(a)} > s_{tol} \tag{9}$$

$$y^{(a)} = y^{(i)} \tag{10}$$

$$\mathbf{r}\left(\mathbf{u}^{(a)}\left[\overline{\mathcal{F}},:\right];k,\boldsymbol{\mu}^{(a)},\boldsymbol{\mu}_{\square}^{(a)}\right)[\mathcal{F}] = 0,\ k = 1,\ldots n_t \tag{11}$$

$$\mathbf{u}^{(a)}[j,:] = \mathbf{u}^{(i)}[j,:] \quad \forall\, j \in \overline{\mathcal{F}}\backslash\mathcal{F} \tag{12}$$

$$p = \ell\left(\mathbf{V}^{(i)}\boldsymbol{\gamma}^{(a)}\right) \tag{13}$$

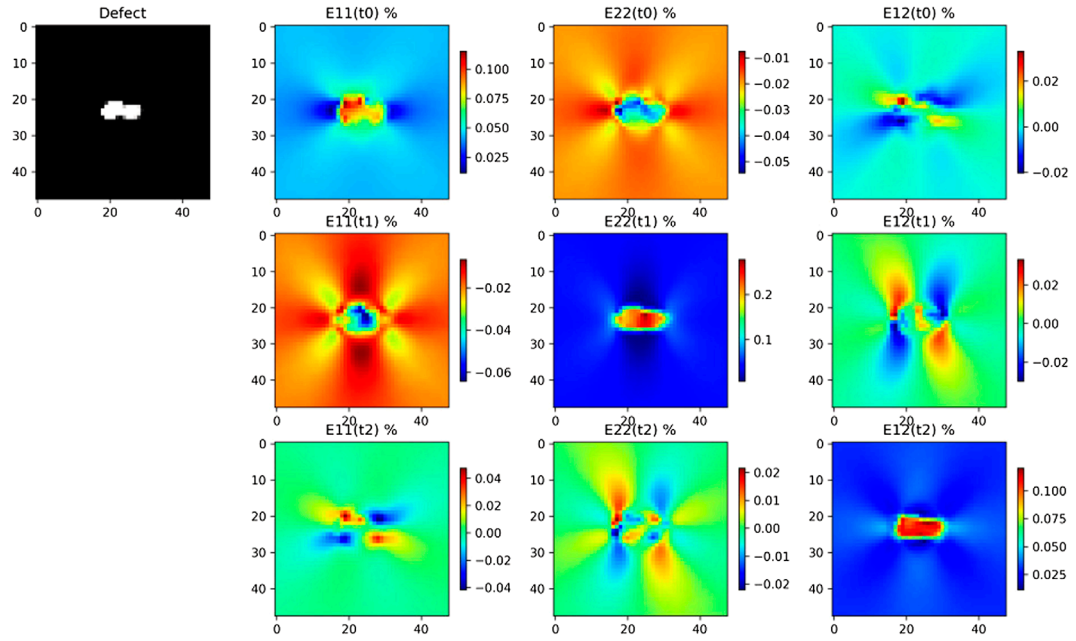

FIGURE 3
Multimodal data for digital twin $i = 1$, on the top left $\mu_{\square}^{(1)}$; on the right, the 3 components of the strain tensor $\varepsilon(\mathbf{u}^{(1)})$ at 3 time instants $t0$, $t1$, $t2$.

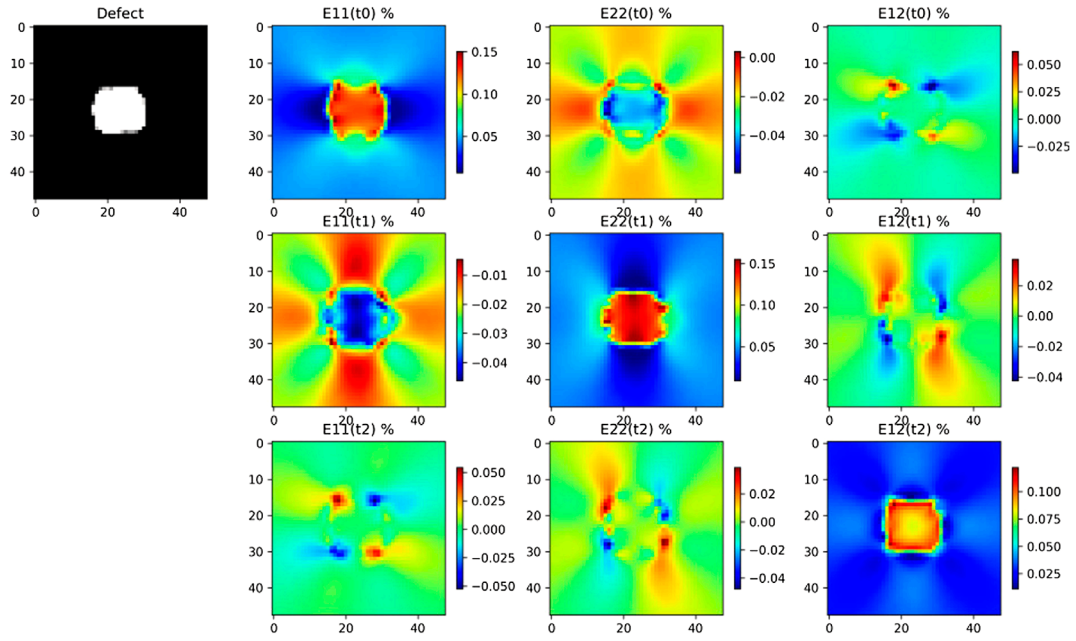

FIGURE 4
Multimodal data for digital twin $i = 2$, on the top left $\mu_{\square}^{(2)}$; on the right, the 3 components of the strain tensor $\varepsilon(\mathbf{u}^{(2)})$ at 3 time instants $t0$, $t1$, $t2$.

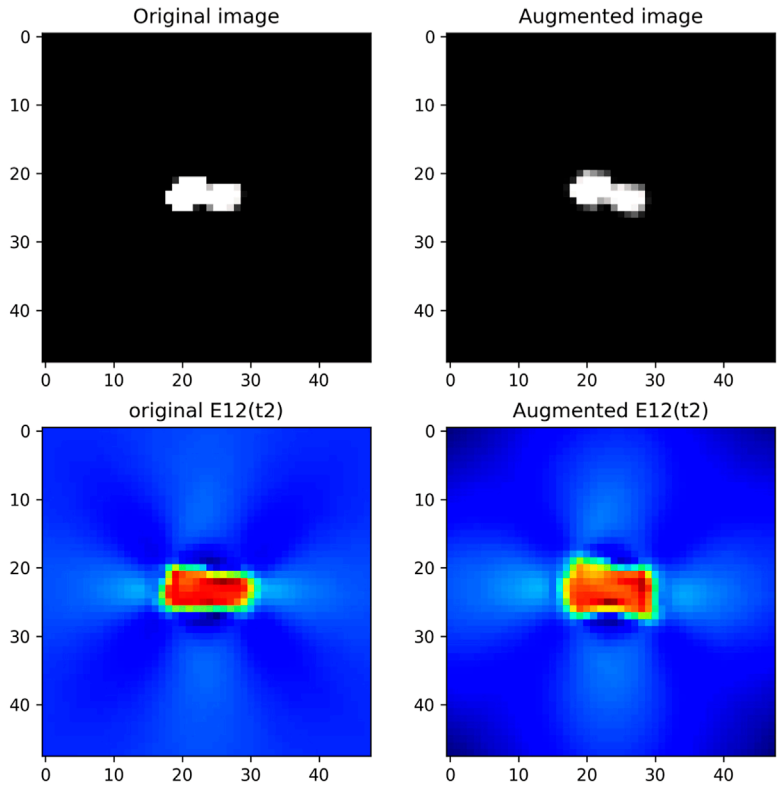

FIGURE 5
Example of augmented multimodal data for digital twin $i = 1$: on the left, original data; on the right, augmented data; in the first row, $\mu_{\square}$; and in the second row, the shear strain at time $t2$.

where $\|\cdot\|_F$ is the Frobenius norm. Here, we assume that $\mathbf{V}^{(i)}[\overline{\mathcal{F}},:]$ is full column rank. The above computation of the reduced coordinates $\boldsymbol{\gamma}^{(a)}$ is known as the Gappy POD (Everson and Sirovich, 1995). Again, we assume that $\mathbf{u}$ is supported by the mesh of the i^{th} DT, so the evaluation of $\eta^{(a)}$ makes sense. **Equation 9** checks the ROM-similarity of the augmented data. **Equation 12** in finite element modeling is termed a Dirichlet boundary condition. Other interesting boundary conditions are proposed in Hilth et al. (2019) for hyper-reduction. s_{tol} is a hyper-parameter of the augmentation scheme.

Remark: It is possible to update the vector bundle after the data augmentation. The accuracy of the LROM that we can compute after data augmentation is not directly related to the threshold used here for the ROM-similarity. For instance, after data augmentation, we can aggregate in a matrix $\widehat{\mathbf{u}}^{(i)}$ the primal variable having the same label, for both original and augmented data. Therefore, a truncated singular value decomposition of $\widehat{\mathbf{u}}^{(i)}$ can give a very accurate reduced basis for the category of data $y = y^{(i)}$.

3 Results and discussion

3.1 Modeling the local mechanical response of voids in an elastic material

In a previous work (Ryckelynck et al., 2020), we published 2D DTs of voids in elastic bodies. In this example, we augment the data around two DTs from this previous work. The augmentation ratio is 100 for each DT, prior to fulfilling the constraints for the construction of $\mathcal{D}_{aug}$. Here, mechanical variables are structured data similar to digital images (Launay et al., 2021a). But not all data augmentation techniques proposed in the literature for digital images are relevant for reduced digital twinning. For instance, rotation and shear transformations must be consistent for both geometrical and mechanical data. In this example, the domain Ω is a square (on the left in **Figure 2**). The manufacturing process creates voids, and each instance of the void is placed in the center of Ω. The mesh is specific to each DT in order to capture the geometry of the void (on the right of **Figure 2**). In this mesh, a phase field represents the void: if $\phi(\boldsymbol{\xi}) > 0$,

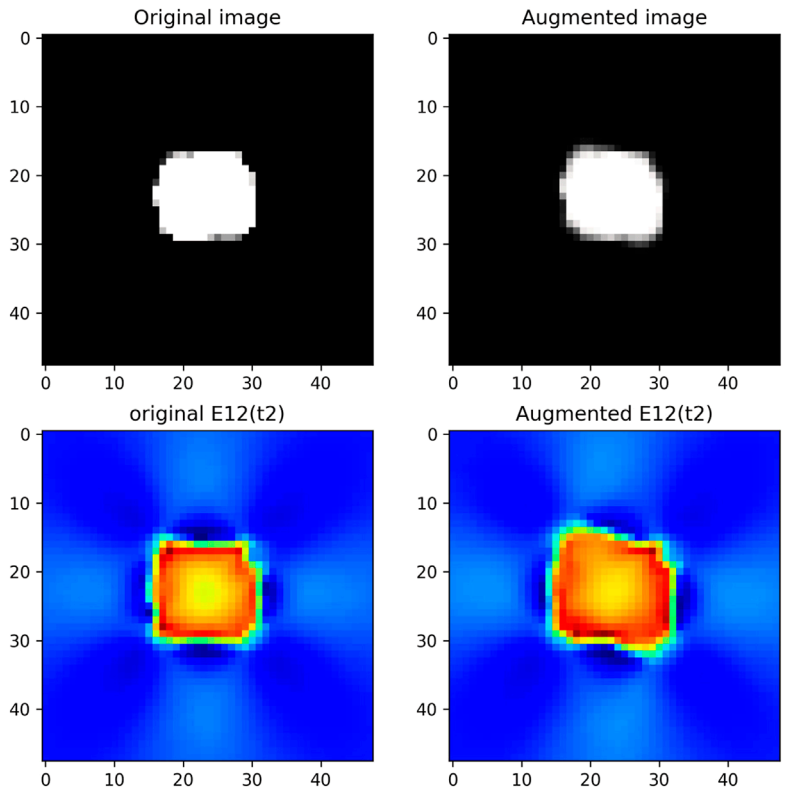

FIGURE 6
Example of augmented multimodal data for digital twin $i = 2$: on the left, original data; on the right, augmented data; in the first-row, $\boldsymbol{\mu}_{\square}$; in the second row, the shear strain at time $t2$.

then $\boldsymbol{\xi}$ is in the void; else, $\boldsymbol{\xi}$ is outside the void ($1 \geq \phi \geq 0$). In mechanical equations of elasticity for digital twining, the Young modulus reads $E(\boldsymbol{\xi}) = E_o(1 - \phi(\boldsymbol{\xi}))$, where E_o is the Young modulus in the bulk. In this simple example, we restrict our attention to the prediction of shear strains around each void, under a macroscopic shear strain. The reduced domain is shown at the bottom of Figure 2. Here, the cardinal of $\mathcal{F}$ is only 30% smaller than $\mathcal{N}$. Simulation errors of shear strains are about 18% when solving Equations 11 and 12 instead of Equation 1. The mechanical problem has 3 time steps $t0$, $t1$, and $t2$. In the first time step, longitudinal traction is imposed to the boundaries of Ω. In the second time step, it is transverse traction. In the last time step, we impose a shear strain at the boundaries of Ω.

Multimodal data are the 2D image of the void (i.e., the image of ϕ) and the mechanical shear strain in Ω_R. These data are shown in Figure 3 and Figure 4 for digital twins $i = 1$ and $i = 2$, respectively. Augmented image data are obtained by usual shear and rotation of the void image. The resulting phase field is denoted by ϕ'. The mechanical expression of the strain tensor is the symmetric part of the gradient of the displacement field $\mathbf{u}$. This strain tensor is denoted by $\boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}(\mathbf{u}) = (\nabla\mathbf{u} + \nabla\mathbf{u}^T)/2$. The solution of elastic Equations 11 and 12 forecasts the augmented value for the shear strain $\boldsymbol{\varepsilon}(\mathbf{u}')$. We can see in the full order simulations, in Figure 3 and Figure 4, that the mechanical strains are highly dependent on the void morphology.

Two examples of augmented data are shown in Figure 5 and Figure 6. In these examples, it is clear that the mechanical variables do not follow the warping imposed to $\boldsymbol{\mu}_{\square}$, especially far from the void. Local reduced bases have 3 modes. The reduced basis of the first DT is shown in Figure 7. The ROM-similarity of augmented data is shown in Figure 8. Only one set of augmented data in $\widetilde{\mathcal{D}}^{(1)}$ fails to be ROM-similar. They are 28 for the second DT. The blue and red curves are well separated. None of the augmented data for one DT is ROM-similar to the other DT.

3.2 Application on fatigue specimen observed *via* X-ray computer tomography

In this subsection, we consider a mechanical specimen subjected to a mechanical test at high temperature, where as-manufactured geometry may impact the lifetime of the specimen. The DT includes a 3D X-Ray tomography of the specimen

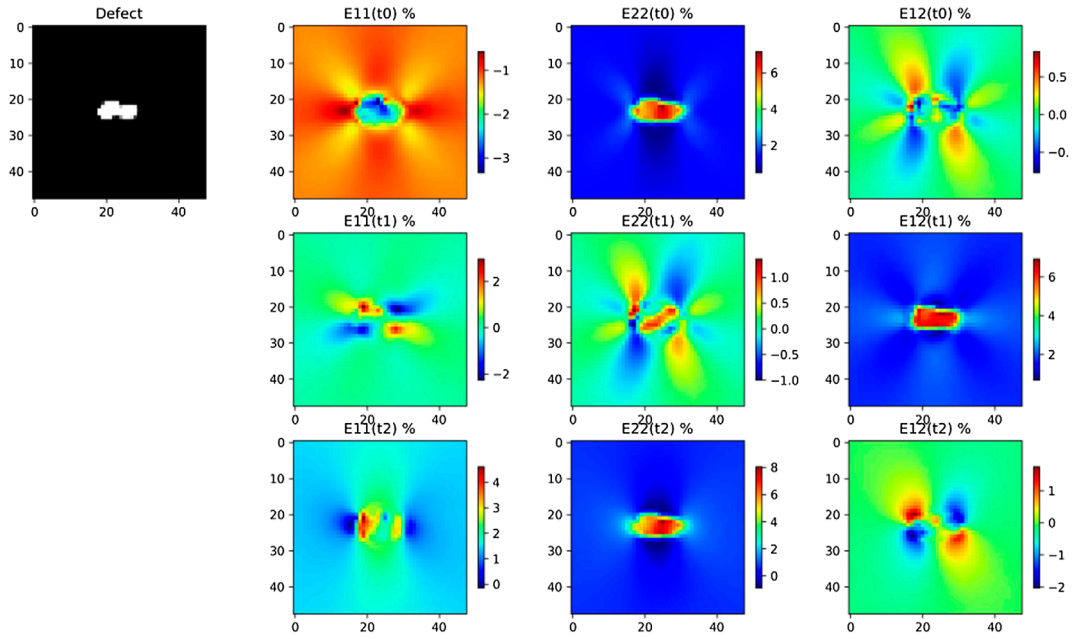

FIGURE 7
Strain components of the LROM for digital twin $i = 1$: on the top left, $\mu_{\square}$; on the right, $\varepsilon\left(\mathbf{V}^{(1)}\right)$. Here, $t0$, $t1$, and $t2$ stand for the indices of the vectors that span the LROM.

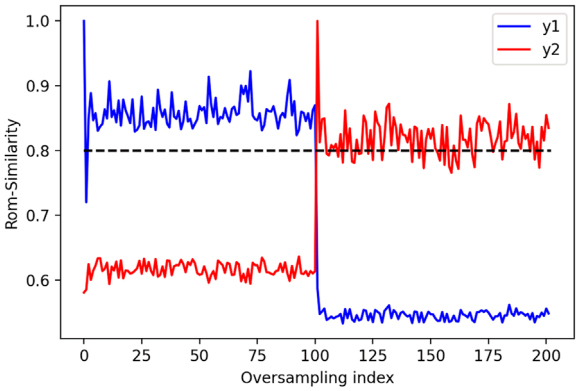

FIGURE 8
ROM-similarity for all instances of augmented data in $\widetilde{\mathcal{D}}^{(i)}$. Augmented data related to digital twin 1 have indexes ranging from 1 to 100, and augmented data related to digital twin 2 have indexes ranging from 101 to 200. The dashed line is the constrain $\eta^{(i)}(\mathbf{u}\prime) < \alpha\, \epsilon_{tol}$.

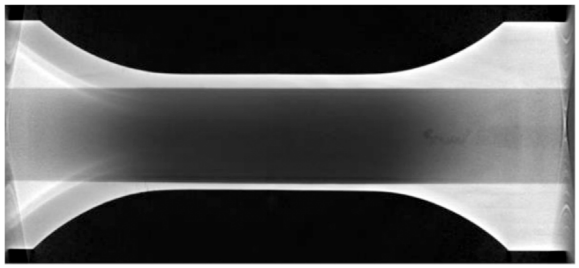

FIGURE 9
X-Ray computer tomography image of the mechanical specimen.

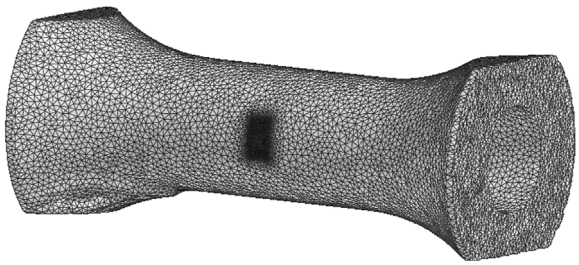

FIGURE 10
3D finite element mesh based on the X-Ray computer tomography image.

(Figure 9), the related finite element mesh (Figure 10), and the finite element predictions of thermo-mechanical variables that are modified at several time steps during the test (on the pruned mesh in Figure 13). The problem of data scarcity is particularly pronounced for this kind of DT. In this example, $\mathcal{N} > 3.10^5$, $m = 20$, and the target reduced dimension of LROM is set up to

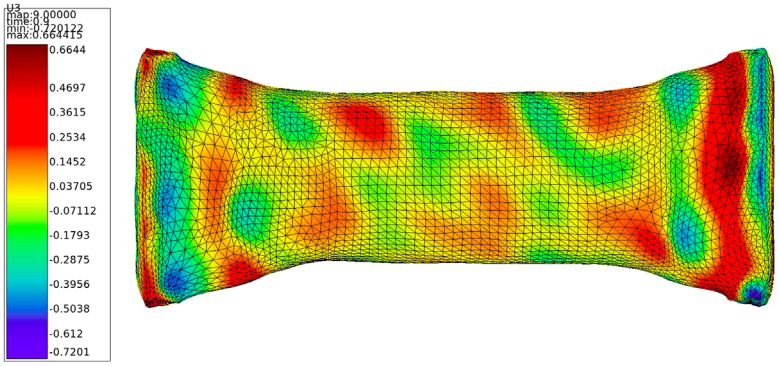

FIGURE 11
Example of node displacements along axial direction for mesh morphing according to augmented data for geometrical variations $\Delta\mu_{\square}$.

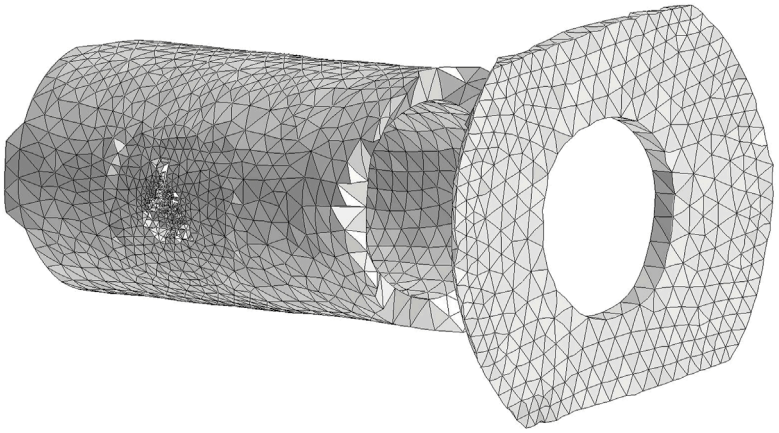

FIGURE 12
Finite element mesh restrained to Ω_R ($\mathcal{N}/\mathrm{card}(\overline{\mathcal{F}}) = 2$). Most removed elements are located around the hole.

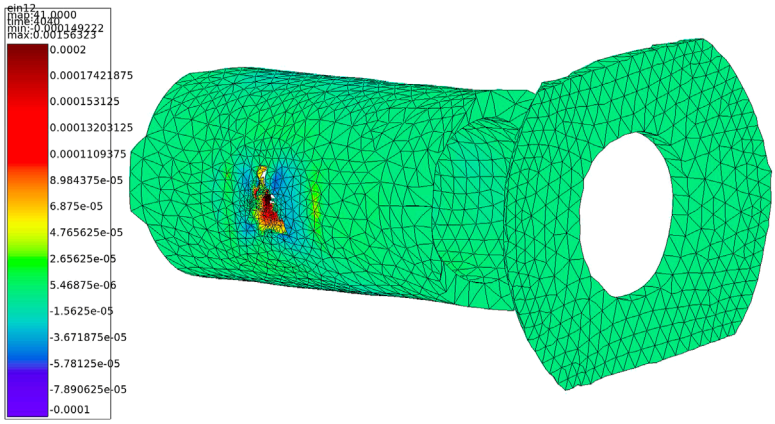

FIGURE 13
Shear strain forecast by the finite element model, in Ω_R, at the maximum mechanical loading of the specimen.

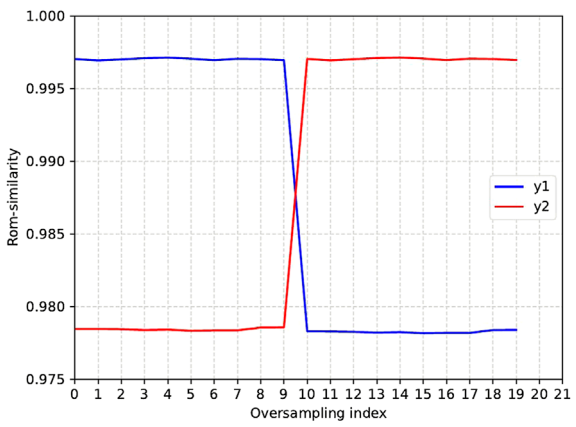

FIGURE 14
ROM-similarity for all instances of augmented data in $\widetilde{\mathcal{D}}^{(i)}$. Augmented data related to digital twin 1 have indexes ranging from 0 to 9, and augmented data related to digital twin 2 have indexes ranging from 10 to 19.

$N = 6$ in order to have $s_{tol} = 0.99$. In this example, $\boldsymbol{\mu}$ is a vector of 3 angles related to the single crystal CMSX4-PLUS of the specimen as presented in Aublet et al. (2022). This parameter is augmented randomly using angle variations below 12^o. Usual image warping techniques are not relevant to the 3D tomographic image of DT. Hence, a random mesh morphing has been applied to the mesh of the DT in order to augment observational data $\boldsymbol{\mu}_{\square}$. The geometrical variations were generated around a DT^i using a Gaussian perturbation on geometrical parameters with the following model. For a parameter q, the new value p' was randomly chosen in the vicinity of q with a common log-normal $\mathcal{X}$ law centered on q with a standard deviation of $0.2 * q$ such as $p' \in \mathcal{X}(q, 0.2 * q)$. We generated 10 meshes around each DT, and 10 cycles of triangular thermo-mechanical loading were applied. An example of node displacement for this morphing step is shown in **Figure 11**. Most finite elements removed for the data pruning are located around a cooling hole in the specimen (**Figure 12**). To ensure the consistency of HROM in **Equation 11** under Neumann boundary conditions, some sets have to be preserved during the pruning. Card($\overline{F}$) is twice as small as $\mathcal{N}$, so the memory saving factor is 2. The prediction error of finite element simulation using the pruned mesh is 10% in Ω_R. Data augmentation has been performed for two DTs $X^{(1)}$ and $X^{(2)}$. After augmentation, the ROM-similarity is computed twice for all augmented data. This enables us to check the labels of augmented data. The original DTs are quite ROM-similar, so values of s are close to 0.99. They are reported in **Figure 14**. Labels of augmented data are consistent with the ROM-similarities. However, all ROM-similarities are very high in this example.

Here, we consider a time-dependent problem due to plasticity computed with a crystal plasticity model developed by Méric et al. (1991). This kind of problem implies a temporal analysis with an initial important spread on the domain (low values with importation gradients) the first time, and then, a localization due to a mechanical defect such as a hole as shown in **Figure 10**. This localization implies local variations of displacement fields around the defect which needs a higher value of N to catch the information in LROM. As we performed a truncature with $N = 6$, we cannot gather the local information on plasticity. The development of ROM-similarity criteria is still a work in progress in plasticity. Here, $\mathbf{V}(y^{(1)})$ and $\mathbf{V}(y^{(2)})$ seem pretty similar, hence the close ROM-similarities reported in **Figure 14**.

4 Conclusion

The proposed data augmentation method is very versatile for multimodal data that includes simulated data in the framework of image-based digital twinning. It requires access to the finite element model used for digital twinning. The method assumes that the latent space hidden in the multimodal data is piecewise linear, so the nonlinear model reduction uses local reduced bases. The validity domains of these reduced bases enable the definition of similarities, termed ROM-similarity, between multimodal data. The pruning technique, used to limit memory requirement for high-dimensional augmented data, has a hyper-parameter Ω_R. The larger the domain Ω_R, the more accurate the evaluation of ROM-similarities. The smaller the Ω_R, the higher the memory savings. The development of ROM-similarity criteria is still a work in progress in plasticity.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

Author contributions

AA performed the digital twinnings for the thermo-mechanical specimens and made the X-Ray computed tomographies. FN did all image-based finite element meshes. HP brought his help for the elasto-viscoplastic simulations, and edited and commented on the initial manuscript. DR proposed the augmentation scheme, and he applied it to the numerical results on the 2D example. AA applied the augmentation scheme to the thermo-mechanical specimens.

Funding

AA's PhD thesis is financially supported by SAFRAN.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aublet, A., Rambaudon, M., N'Guyen, F., Ryckelynck, D., Remacha, C., Cariou, R., et al. (2022). Mechanical fatigue testing under thermal gradient and manufacturing variabilities in nickel-based superalloy parts with air-cooling holes. *Exp. Mech.* doi:10.1007/s11340-022-00868-0

Bellinger, N., Tuegel, E. J., Ingraffea, A. R., Eason, T. G., and Spottswood, S. M. (2011). Reengineering aircraft structural life prediction using a digital twin. *Int. J. Aerosp. Eng.* 1, 14. doi:10.1155/2011/154798

Daniel, T., Casenave, F., Akkari, N., and Ryckelynck, D. (2021). Data augmentation and feature selection for automatic model recommendation in computational physics. *Math. Comput. Appl.* 26, 17. doi:10.3390/mca26010017

Daniel, T., Casenave, F., Akkari, N., and Ryckelynck, D. (2020). Model order reduction assisted by deep neural networks (ROM-net). *Adv. Model. Simul. Eng. Sci.* 7, 16. doi:10.1186/s40323-020-00153-6

Everson, R., and Sirovich, L. (1995). Karhunen-Loeve procedure for gappy data. *J. Opt. Soc. Am. A* 12, 1657. doi:10.1364/JOSAA.12.001657

Glaessgen, E., and Stargel, D. (2012). "The digital twin paradigm for future nasa and u.s. air force vehicles," in Structures, Structural Dynamics, and Materials Conference (Aerospace Research Central). doi:10.2514/6.2012-1818

Hilth, W., Ryckelynck, D., and Menet, C. (2019). Data pruning of tomographic data for the calibration of strain localization models. *Math. Comput. Appl.* 24, 18. doi:10.3390/mca24010018

Launay, H., Ryckelynck, D., Lacourt, L., Besson, J., Mondon, A., and Willot, F. (2021a). Deep multimodal autoencoder for crack criticality assessment. *Int. J. Numer. Methods Eng.* 1, 1456–1480. doi:10.1002/nme.6905

Launay, H., Willot, F., Ryckelynck, D., and Besson, J. (2021b). Mechanical assessment of defects in welded joints: Morphological classification and data augmentation. *J. Math. Ind.* 11, 18. doi:10.1186/s13362-021-00114-7

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi:10.1109/5.726791

Méric, L., Poubanne, P., and Cailletaud, G. (1991). Single crystal modeling for structural calculations: Part 1 — model presentation. *J. Eng. Mat. Technol.* 113, 162–170. doi:10.1115/1.2903374

Ryckelynck, D., Gallimard, L., and Jules, S. (2015). Estimation of the validity domain of hyper-reduction approximations in generalized standard elastoviscoplasticity. *Adv. Model. Simul. Eng. Sci.* 2, 6. doi:10.1186/s40323-015-0027-7

Ryckelynck, D., Goessel, T., and Nguyen, F. (2020). Mechanical dissimilarity of defects in welded joints via Grassmann manifold and machine learning. *Preprint.*

Seon, G., Nikishkov, Y., Makeev, A., and Ferguson, L. (2020). Towards a digital twin for mitigating void formation during debulking of autoclave composite parts. *Eng. Fract. Mech.* 225, 106792. doi:10.1016/j.engfracmech.2019.106792

Shorten, C., and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *J. Big Data* 6, 60. doi:10.1186/s40537-019-0197-0

Zhang, D., Lu, L., Guo, L., and Karniadakis, G. E. (2019). Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems. *J. Comput. Phys.* 397, 108850. doi:10.1016/j.jcp.2019.07.048