VolRAFT: Volumetric Optical Flow Network for Digital Volume Correlation of Synchrotron Radiation-based Micro-CT Images of Bone-Implant Interfaces

Tak Ming Wong[1,2*]   Julian Moosmann[2]   Berit Zeller-Plumhoff[1*]

[1]Institute of Metallic Biomaterials, Helmholtz-Zentrum Hereon, Germany
[2]Institute of Materials Physics, Helmholtz-Zentrum Hereon, Germany

{tak.wong, julian.moosmann, berit.zeller-plumhoff}@hereon.de

Abstract

In materials science research, digital volume correlation (DVC) analysis is commonly used to track deformations and strains to elucidate morphology-function relationships. Optical flow-based DVC is particularly popular because of its robustness to estimate the correlation as a dense deformation vector. Recently, computer vision researchers showed that network-based optical flow approaches can outperform classical iterative optical flow approaches. In this paper, we propose a supervised machine learning approach for digital volume correlation, VolRAFT, that estimates the 3D displacement vector between the reference volume and the deformed volume. The proposed approach extends the state-of-the-art network-based optical flow method, RAFT, from 2D images to 3D volumes such that it predicts the volumetric displacement vector from the input volume pairs. Experiments show that the proposed network performs well in estimating different displacement fields when compared to cutting-edge iterative DVC methods for bone-implant materials based on high resolution synchrotron-radiation micro-computed tomography imaging data.

1. Introduction

Digital volume correlation (DVC) analysis is a highly important tool in materials science to track deformations and strains. Thus, it can be used to evaluate a material's functional behaviour from *in situ* imaging and determine morphology-function relationships. For example, DVC based on synchrotron radiation-based micro-CT (SRµCT) and laboratory µCT data has been used for a long time to study deformations in musculoskeletal tissues [9], as well as materials such as fibre-reinforced composites [16]. Most of the existing DVC analysis methods can be categorised into local or global methods [8, 9]: local DVC methods [2, 18]
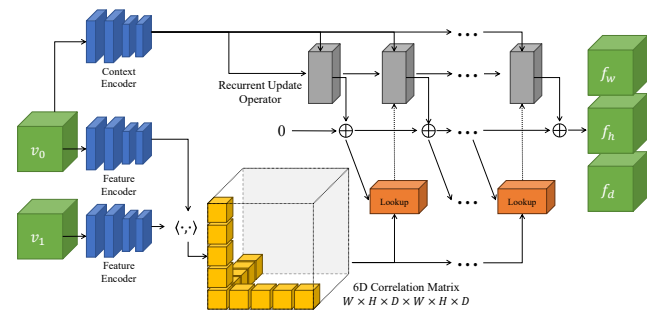

Figure 1. VolRAFT consists of three major parts: feature encoders that extract per-voxel feature vectors from both volumes; dot product of feature vectors that constructs a 6D correlation matrix (in the dimension of $W \times H \times D \times W \times H \times D$); recurrent 3D update operators which update the displacement field f. The update operators predicts the 3D displacement field in its dimensions of width (f_w), height (f_h) and depth (f_d) respectively.

estimate the displacement field as a local neighbourhood block-matching problem; global DVC methods [7, 11] formulate the DVC problem as a global variational problem.

Interestingly, the trend of methodological development to move from local to global methods is highly similar to the development of optical flow method in computer vision area, where optical flow is defined as the pixel motion of adjacent frames by the dense displacement field [36]. This similarity of development trends is because the objective of both optical flow and DVC methods is to estimate the dense displacement vectors between images (i.e. the deformation vectors for volumes), which makes optical flow-based DVC [7, 11, 15, 23] particularly interesting because of the robustness shown by optical flow methods.

However, classical iterative optical flow methods often require hand-crafting and fine-tuning by experts [31], which creates a bottleneck in their application for domain experts such as materials scientists, as individual volumetric data can have a large variety of tested samples and experimental setups. Thus, machine learning-based optical flow methods,

---

*Corresponding authors

which are also categorized as data-driven approaches [36], are of high interest as they learn the prior knowledge instead of requiring expert-handcrafting. Recently, researchers showed that learning-based approaches can outperform classical iterative optical flow approaches [36].

Practically, modern machine learning approaches rely heavily on the computing power of GPUs, which may represent a limitation based on the GPU memory size. To apply machine learning based optical flow methods to DVC analysis, the exponential increase of memory requirement brought by the increment of data dimensionality from 2D images to 3D volumes is one of the major challenges, especially for high resolution volumetric data such as the tomography data by SRμCT imaging.

In this paper, we apply machine learning-based optical flow methods to the DVC analysis of *in situ* SRμCT data of loading of bone-implant specimen. We extend one of the state-of-the-art optical flow networks, the Recurrent All-Pairs Field Transforms (RAFT) [31], from 2D image pairs to 3D volume pairs as a Volumetric RAFT (*VolRAFT*) approach, which estimates the 3D displacement vectors between the reference volume and the deformed volume. In detail, our contribution entails:

- The proposal of a supervised machine learning approach for digital volume correlation, *VolRAFT*, that estimates the 3D displacement fields between the reference volume and the deformed volume.

- An extension of the optical flow neural network approach, Recurrent All-Pairs Field Transforms (RAFT) [31], from 2D images to 3D volumes, which includes the volumetric input and output tensors, network layers and the 6D correlation matrix.

- The generation of synthetic displacement fields and their application to the measured tomographic volume by volume warping, and usage of the deformed volume to serve as the training, validation and testing datasets.

- A comparison to cutting-edge iterative methods, for bone-implant loading scenarios based on SRμCT image data, which reveals that the proposed VolRAFT approach can achieve a better performance.

2. Related work

Digital Volume Correlation (DVC)   DVC is a commonly used method for materials analysis to study the deformation and strain for materials experiments. Hussein *et al*. [18] formulated DVC as a maximum likelihood estimation and solved it by the Gauss-Newton method, which is a gradient-based iterative method. Bar-Kochba *et al*. [2] proposed Fast Iterative Digital Volume Correlation Method (FIDVC) to estimate the dense displacement field in local windows. Dos Santos Rolo [11] expressed the DVC analysis as a global 3D variational optical flow problem

and then solved it by the Brox's variational optical flow method [6]. Hermann and Werner [15] and Nogatz *et al*. [23] also extended the Brox's optical flow method to compute the correlation vector of 3D CT images. Recently, Bruns *et al*. [7] proposed and implemented an iterative DVC method based on the optical flow method in [11] for the study of implant materials that was similarly applied successfully to other materials [24–26, 32]. Other common methods also included the commercial software DaVis (LaVision GmbH, Göttingen, Germany), the service BoneDVC (Insigneo Institute, Sheffield, UK), CCPi iDVC app (Collaborative Computational Project in Tomographic Imaging, UK [10, 22]) and TomoWarp2 [34]. For further details, since the pioneering review by Bay in [3], Buljac *et al*. [8] provided a review from the perspective of solid mechanics, while Dall'Ara and Tozzi[9] conducted a survey on the application of DVC methods in biological tissues.

Using machine learning for DVC is a relatively new direction in domain research fields, although it has shown improvements in efficiency by reducing the computational complexity [9]. Shen *et al*. [27] trained several common deep CNN architectures, such as AlexNet and ResNet, in supervised and transfer learning manners for DVC analysis of *in situ* SRμCT experimental data of bone images. Duan and Huang [13] proposed a deep learning based approach, DVC-Net, to train convolutional neural networks (CNNs) directly predicting the displacement vectors from the volumetric intensities for simulated data and volumetric images captured by a laser scanning confocal microscope. Wang *et al*. [35] proposed a deep learning based approach, StrainNet-3D, for real-time DVC based on optical flow CNN using simulated data. However, most of these CNNs basically only enlarged commonly used network architectures with hyperparameter fine-tuning. It could be more beneficial to explore state-of-the-art networks designed for optical flow with a proven improvement beyond CNNs.

Machine Learning based Optical Flow   Machine learning based optical flow method is a well-studied problem in computer vision and pattern recognition area, where researchers provided comprehensive reviews in [14, 33, 36]. In this section, we focus on popular neural network approaches based on supervised training methods. Modern optical flow networks started to emerge from FlowNet [12] and its later extension FlowNet2 [19], which consisted of a fully convolutional neural network (actually a U-Net) and a combination of multiple U-Nets. On top of FlowNet, PWC-Net [30] was proposed with respect to its feature *P*yramid, *W*arping and *C*ost volume, while the authors showed that it was more efficient and faster than FlowNet2. After PWC-Net, Teed and Deng proposed RAFT [31], which was later considered as one of the most influential network based approach [29], as RAFT demonstrated an outstanding performance in terms of robustness and memory ef-

ficiency. Most of the well-known optical flow in recent years were generally developed based on RAFT, namely BRAFT [20], CRAFT [29], FlowFormer [17] and Flow-Former++ [28], *etc.* Even if researchers introduced transformers to the RAFT architecture [17, 28], the original RAFT approach based on convolutional layers is still considered one of the most renowned optical flow methods because of its simplicity and efficiency.

Recurrent All-Pairs Field Transforms (RAFT)   As our approach extends the approach of RAFT [31], we discuss it in more detail. The RAFT approach simplifies optical flow estimation into a three-step process: (1) a pair of feature encoders transform each pixel of the images into a pair of latent feature space vectors. (2) a correlation layer computes the similarity between all feature pairs, constructing a 4D correlation volume. A pyramid of 4D correlation volumes is constructed by applying average pooling to this 4D correlation volume. This correlation pyramid further extracts a sequence of feature maps in lower dimensions. (3) the recurrent update operator predicts values from the pyramid of correlation feature maps and iteratively refines the flow field from an initial zero state.

According to the authors in [31], RAFT is inspired by classic optimization techniques, but with a critical difference: both the feature extraction and the correlation estimation are learned from the data, allowing RAFT to adapt to complex and varied motion patterns more effectively than traditional, handcrafted methods. Therefore, RAFT seems ideal for the extension towards volumetric image data, as we outline in the following.

To the best of our knowledge, the proposed approach is the first work to apply cutting-edge machine-learning-based optical flow approach for DVC analysis for bone-implant specimen using high resolution SRµCT imaging data.

3. The proposed VolRAFT method

Given a pair of volumes reconstructed by CT reconstruction methods, we denote the reference volumetric intensity as $v_0(\mathbf{x})$ and the deformed volumetric intensity as $v_1(\mathbf{x})$ at each spatial position $\mathbf{x}$. DVC is formulated as estimating a dense displacement field f, such that the displacement field maps each voxel in v_0 to its corresponding coordinates in v_1, resulting in $v_1(\mathbf{x}) = v_0(\mathbf{x} + \mathrm{f}(\mathbf{x}))$.

Fig. 1 illustrates the proposed approach, which consists of three major parts: feature extraction by encoders, a 6D correlation matrix, and the 3D recurrent update operators.

Encoders   To extract features from the input volume pairs, two feature encoders and one context encoder are built by convolutional layers (illustrated as blue blocks in Fig. 1). Here, we denote the feature encoders for the reference volume as $g_\theta(v_0) \in \mathbb{R}^{W \times H \times D \times C}$ and for the deformed volume $g_\theta(v_1) \in \mathbb{R}^{W \times H \times D \times C}$, where W, H, D, C

are dimensions of width, height, depth and channels in the feature space respectively, and θ denotes the network parameters. Note that the context and the feature encoders g_θ are almost identical in their architectures, only except that the feature encoders utilize instance normalization while the context encoder does not use any normalization layer.

6D correlation matrix   To compute the visual similarity, we construct a correlation matrix (yellow blocks in Fig. 1) by the dot product operation between the extracted feature vectors and the feature encoders. Given the feature vectors $g_\theta(v_0)$ and the feature vector $g_\theta(v_1)$, we formulate the 6D correlation matrix M as:

$$M\left(g_\theta(v_0), g_\theta(v_1)\right) \in \mathbb{R}^{W \times H \times D \times W \times H \times D}$$

$$\text{where} \quad M_{ijklmn} = \sum_h g_\theta(v_0)_{ijkh} \cdot g_\theta(v_1)_{lmnh} \quad (1)$$

We further construct a correlation pyramid M^L by average pooling the last three dimensions of M by the stride of 2^L, where L denotes the level of pyramid. This multi-scale pyramid $M^L \in \mathbb{R}^{W \times H \times D \times W/2^L \times H/2^L \times D/2^L}$ allows the network to estimate the large and small displacement and to recover the large displacement of small-sized structures [31].

After the correlation matrix, correlation lookup operators (orange blocks in Fig. 1) generate feature maps by indexing from the correlation pyramid. Similar to RAFT [31], these correlation lookup operators use the local neighborhood, which are grids around the projected position on the deformed volume, to index the correlation volumes for each pyramid level. These grids are interpolated by trilinear sampling and are then concatenated to form a feature map.

3D recurrent update operators   Similar to RAFT, the 3D recurrent update operators (grey blocks in Fig. 1) predict a sequence of displacement fields $\{\mathrm{f}_1, ..., \mathrm{f}_N\}$ for each iteration, starting from a hidden initialized state $\mathrm{f}_0 = 0$. At the k-th iteration, this update operator estimates an update of the displacement $\Delta \mathrm{f}$ such that $\mathrm{f}_{k+1} = \mathrm{f}_k + \Delta \mathrm{f}$.

The correlation features from the lookup operator are further extracted by two convolutional layers. Another two convolutional layers are applied to extract the displacement features from the current displacement field f_k. In addition to the direct extracted features from the context encoder, these three features - correlation, displacement, and context - are concatenated and served as the input feature map x_t for the recurrent update operator.

The recurrent update operator is based on a Gated Recurrent Unit (GRU), where its core component is a gated activation unit with 3D convolutional layers:

$$\begin{aligned} z_k &= \text{sigmoid}\left(\text{Conv}_{3 \times 3 \times 3}\left([h_{k-1}, x_k]; \theta\right)\right) \\ r_k &= \text{sigmoid}\left(\text{Conv}_{3 \times 3 \times 3}\left([h_{k-1}, x_k]; \theta\right)\right) \\ q_k &= \text{tanh}\left(\text{Conv}_{3 \times 3 \times 3}\left([r_k \cdot h_{k-1}, x_k]; \theta\right)\right) \\ h_k &= (1 - z_k) \cdot h_{k-1} + z_k \cdot q_k \end{aligned} \quad (2)$$

where $[\cdot,\cdot]$ denotes the concatenating operation and h_k is the hidden state predicted by GRU at the k-th iteration.

The output feature h_k from the GRU block passes through two convolutional layers to predict the displacement update at a lower dimension. Finally, we use trilinear interpolation to upsample the predicted displacement update Δf to the desired dimension $W \times H \times D$.

Supervised loss with a foreground mask   When considering DVC from (SR)µCT, generally the interested material subjected to deformation is placed within a controlled environment and sample holder. These may be visible within the reconstructed image data. Therefore, the DVC analysis should be limited to the sample region itself, for example by creating a foreground segmentation mask [7]. Hence, regarding the loss function, we apply a binary volumetric mask $v_m \in \{0,1\}^{W \times H \times D}$, which separates the background and foreground voxels (*e.g.* in bone-implant samples, the foreground structures include both the screw and the implant), respectively. We train the network using supervised learning approach with the mini-batch **b** to minimize the ℓ^1-distance between the predicted and the ground truth for all of the predicted displacements $\{f_1, ..., f_N\}$:

$$\min_\theta \sum_{\mathbf{b}} \sum_{k=1}^{N} \gamma^{N-k} \cdot v_m \cdot \left\| f_{gt}^{\mathbf{b}} - f_k^{\mathbf{b}}(\theta) \right\|_1 \qquad (3)$$

where γ denotes the exponentially increasing weights and f_{gt} is the ground truth displacement.

Patch-based training and inference   During the adaptation of RAFT from 2D image pairs to 3D volumetric pairs, we discovered that it is very difficult to have sufficient GPU memory to store and train the full volumes, the ground-truth displacement fields, and the network parameters in a single GPU device. Theoretically, it is possible to split all tensors into multiple GPU devices for computation. However, in common practise of material science research, the size of the high-resolution imaging data such as the volumes of synchrotron radation tomography can easily scale up in the third-order of size, *e.g.* the SRµCT volumes in this paper are $1280 \times 1280 \times 960$ voxels. Therefore, we split the reference volume v_0, the deformed volume v_1, the volume mask v_m, the ground truth f_{gt} and the predicted displacement fields f into patches with the size factor of $1/16$ (*i.e.* the full volume with a size of $1280 \times 1280 \times 960$ is split into patches with a size of $80 \times 80 \times 60$). Moreover, we use the half-size overlapping stride ($40 \times 40 \times 30$) and only include patches that contain foreground voxels in its volume mask. We train VolRAFT based on these patch pairs.

As the network is trained using patch pairs, it predicts the displacement field in the dimension of patch size (*i.e.* $80 \times 80 \times 60$ in our case). Hence, we utilize a 4-steps patched-base method for the inference:

- Roll the sampling window by the stride size of $1/7$ (*e.g.* $11 \times 11 \times 8$ in our case) for each direction.

- Generate a 3D Gaussian distribution matrix in the dimension of the patch size as the weights of prediction.
- Multiply and accumulate the Gaussian weights to the predicted displacement, and repeat step 1.
- Normalize the overall displacement field by dividing the weighted sum of fields by the sum of weights.

We have also tried approaches such as median/averaging filtering and Hamming windows blending, while empirically this method obtains the best inference result.

4. Implementation and experiments

In this section, we provide more details about the implementation, experimental setup, and evaluation methods. VolRAFT is implemented on PyTorch (version 2.1.0) using NVIDIA Tesla V100 with 32GB GPU memory. The source code and trained network are available on GitHub `https://github.com/hereon-mbs/VolRAFT`.

Volume normalization   In computer vision area, researchers usually assume the data are normalized vectors in a range of $[0,1]$ for 8-byte RGB images (*e.g.* as RAFT approach assumed). However, in (32-bit) images reconstructed from µCT imaging, this assumption is often invalid as the data contains physical information, *e.g.* the range of SRµCT data related to the measured attenuation of X-ray radiation. Therefore, at the beginning of VolRAFT model, we normalize the reference volume v_0 and the deformed volume v_1 before applying the encoders. We find the minimum value $v_{\min} = \min(v_0 \cdot v_m, v_1 \cdot v_m)$ and the maximum value $v_{\max} = \max(v_0 \cdot v_m, v_1 \cdot v_m)$ across both volumes at foreground positions. By computing $\bar{v}_{\{0,1\}} = (v_{\{0,1\}} - v_{\min})/(v_{\max} - v_{\min})$, the normalized volumes $\bar{v}_0, \bar{v}_1$ are then passed to the network encoders.

Hyperparameters   To demonstrate that our method is a native extension from RAFT without further hyperparameter optimization, we essentially reuse hyperparameters from the RAFT approach [31] using the small model RAFT-S, in which the authors showed a significantly lower number of network parameters. In Tab. 1, we compare the main hyperparameters and the number of learnable parameters between the RAFT-S model and VolRAFT. As the hyperparameters of the original RAFT model were optimized for each 2D optical flow dataset individually by the authors [31], the settings reported in the RAFT paper can be biased for a particular use case. Therefore, to have a fair and generalized setting, we take the default setting of hyperparameters from the published source code of RAFT for our use case.

Measured volumes   The image volumes used in this study were obtained by SRµCT imaging at the P05 imaging beamline operated by the Helmholtz-Zentrum Hereon at PETRA III at Deutsches Elektronen-Synchrotron (DESY, Hamburg, Germany). 1200 projections measuring $5120 \times$

Hyperparameter		RAFT-S	VolRAFT
Size of image/volume-patches		512×384	$\mathbf{80 \times 80 \times 60}$
Size of flow/displacement-patches		$2 \times 512 \times 384$	$\mathbf{3 \times 80 \times 80 \times 60}$
Optimizer		AdamW [21]	AdamW [21]
Learning rate		2e-5	2e-5
Number of epochs		100000	**10000**
Maximum range of displacement		400	**24**
(Mini-)Batch size $\mathbf{b}$	Eq. (3)	6	**18**
Recurrent iterations N	Eq. (3)	12	12
Weights of loss γ	Eq. (3)	0.8	0.8
Levels of correlation matrix L		4	4
Gradient-norm clipped		$[-1, 1]$	$[-1, 1]$
Number of learnable parameters		1M	2.95M

Table 1. Hyperparameters and the number of learnable parameters between the RAFT-S model and VolRAFT. The default hyperparameters from the source code of RAFT are taken as reference. Changes are highlighted.
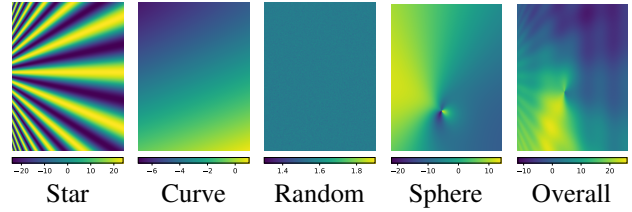

Figure 2. Synthetic displacement fields are generated according to 5 classes: Star, Curve, Random, Sphere and Overall. Slices of displacement field f_w are shown as examples here. As the synthetic displacement fields are augmented by the randomly permuted order of axes, these slices of field can appear in any one of the dimensions (f_w, f_h, f_d) during the generation of synthetic dataset.

3840 pixels were acquired using fly scans with an exposure time of 34 ms per projection. The projection images were processed through the typical CT reconstruction pipeline, *i.e.* flat-field and dark-field correction, binning, ring removal, and tomographic reconstruction, resulting in volumes of $1280 \times 1280 \times 960$ at $5.1\mu m$ voxel resolution. 3D registration was performed to maximize the correlation between the corresponding reference and deformed volumes, and the 3D foreground layers were segmented to define the region of interest on the reference volume.

The measured volumes consisted of rat bone samples containing a screw made of one of four different types of implant materials: magnesium-10wt.%gadolinium (Mg-10Gd), magnesium-5wt.%gadolinium (Mg-5Gd), titanium (Ti) and polyetheretherketone (PEEK). Each implant material was tested after 4, 8, and 12 weeks of healing. During push-out tests, samples were subjected to forces from 5N up to 240N. For more details on the measurement setup and image processing, the readers can refer to [7]. In total, we utilize 39 measured volumes which are used as reference volumes for the generation of synthetic datasets.

Synthetic displacement fields In this paper, we generate 5 different classes of synthetic displacement fields for training and testing (see Fig. 2):

- Star: star field was proposed [4] to access the measurement errors of digital image correlation [4, 5] and DVC [13]. The displacement field is based on synthetic fields undergoing sinusoidal displacements with various amplitudes and spatial frequencies [4].
- Curve: To model single direction deformation, a curved displacement field is generated by 1D curves with an exponential non-linearity term (*i.e.* $f(x) = mx^\alpha + c$, where $1 \le \alpha \le 2$) for all three directions.
- Random: To address the local inhomogeneity of displacement field, a random field with Gaussian distribution is generated.
- Sphere: To simulate various flow dynamics scenarios such as divergence and curl, the sphere displacement

field generates a 3D vector field within a spherical domain based on radial and tangential components.
- Overall: this is a mixture of displacement field over all (Star, Curve, Random and Sphere) displacement fields.

Data augmentation We select 23 different field settings (5 for each single-type class and 3 for the overall field), with each setting being randomly augmented across 39 volumes. These settings, as shown in Fig. 2, undergo random augmentations including rotation by 3D angles, permutation of the axes order, addition of noise to the volume pairs after warping, and generation of random flow parameters (e.g., amplitudes, spatial frequencies, non-linearity, circulating phase, etc.). In total, we generate 897 unique fields for deforming volume pairs, which serve as synthetic fields for training, validation, and testing.

Synthetic and measurement datasets We create the deformed volumes v_1 by warping the synthetic fields to the measured volumes as the reference volume v_0 using the PyTorch function `grid_sample` in the trilinear interpolation mode. Hence, by computation for each reference volume and synthetic fields, we compute 897 (23×39) volume pairs as the total number of datasets.

To ensure the generalization of the model training, we separate these datasets as follows: 828 for training sets, while 20% of the set are randomly split for validation; 69 for testing. We excluded 3 samples (`Mg-5Gd_4w_103L`, `Ti_4w_5R`, `PEEK_4w_5L`) and their corresponding 23 synthetic fields for each sample (*i.e.* training and testing sets are mutually exclusive in samples), so that the generalization of VolRAFT approach can be evaluated.

To verify the generalization of the proposed VolRAFT approach, we also examine the volume pairs, which were experimentally measured for both reference and deformed volume, as the realistic measurement dataset. In this paper, we select the `PEEK_4w_5L` sample using its first and second scan, as well as the `Ti_4w_5R` sample using its first and last scan as reference and deformed volumes, respectively.

Evaluation metrics   In this paper, we investigate the DVC analysis in a more computer-vision-oriented perspective. Instead of measuring the physical displacement (*e.g.* the distance between the reference and deformed structures in μm), we evaluate the result based on the average End-Point-Error (EPE), which is defined as the Euclidean distance [1, 14] between the displacement vector $f = \{f_w, f_h, f_d\}$ and its reference vector $f^{\text{ref}} = \{f_w^{\text{ref}}, f_h^{\text{ref}}, f_d^{\text{ref}}\}$:

$$\text{EPE} = \sqrt{(f_w - f_w^{\text{ref}})^2 + (f_h - f_h^{\text{ref}})^2 + (f_d - f_d^{\text{ref}})^2} \quad (4)$$

This is also because of the difficulty to obtain the ground truth measurement of the physical displacement when examining continuously deforming materials in a high precision modality, *e.g.* the bone-implant materials using SRμCT imaging.

Methods for comparison   In this paper, we compare the VolRAFT approach to the iterative method proposed by Bruns *et al*. [7], which is one of the publicly available state-of-arts methods for high resolution SRμCT imaging data. We apply the open-source code of this *MBS-3D-Optflow* method based on its default settings as the vanilla version, abbreviated MBS-VAN. To have a fair comparison with VolRAFT, we empirically optimize Bruns's method based on one of the testing data sets (*i.e.* the Mg-5Gd_4w_103L sample with the star displacement field) as the optimal version, abbreviated as MBS-OPT.

5. Evaluation

The computation of optical flow in texture-less regions is generally difficult [1]. In this section we thus only evaluate the bone structure but not the texture-less screw structure. We also crop out the sample holder region as it does not contain information of interest for the experiment. We evaluate the 3D displacement fields using the synthetic and the realistic measurement dataset.

Tab. 2 shows the average EPE of all 15 synthetic testing datasets (*i.e.* PEEK_4w_5L, Ti_4w_5R and Mg-5Gd_4w_103L samples for all 5 displacement fields) using the default vanilla setting (MBS-VAN) and the optimal setting (MBS-OPT) of Bruns's method [7]. As the numerical results show, VolRAFT achieves significantly better performance in the Star displacement field than MBS-VAN and MBS-OPT, while it performs in par in Curve, Random, Sphere and Overall fields. For the Ti_4w_5R dataset, VolRAFT appears to yield better results also for the Sphere and Overall displacement fields. Star and Sphere fields are both vector fields that model the blending and the spherical flow dynamics, and the Overall field is a mixture of all classes of linear and rotational fields. Thus, VolRAFT approach generally performs better in displacement fields with strong and significant divergence and curl.

Fig. 3 compares the ground truth displacement fields

Sample	Disp. class	MBS-VAN	MBS-OPT	VolRAFT
PEEK_4w_5L	Star	20.686	24.406	**4.968**
PEEK_4w_5L	Curve	**0.680**	0.714	1.276
PEEK_4w_5L	Random	**0.187**	0.250	0.456
PEEK_4w_5L	Sphere	1.045	1.178	**0.880**
PEEK_4w_5L	Overall	1.195	1.406	**0.882**
*Mg-5Gd_4w_103L	*Star	20.526	20.136	**5.748**
Mg-5Gd_4w_103L	Curve	**0.356**	0.413	0.440
Mg-5Gd_4w_103L	Random	**0.140**	0.152	0.339
Mg-5Gd_4w_103L	Sphere	1.569	2.081	**0.855**
Mg-5Gd_4w_103L	Overall	1.149	1.280	**0.944**
Ti_4w_5R	Star	21.262	19.805	**4.426**
Ti_4w_5R	Curve	**0.923**	0.997	1.270
Ti_4w_5R	Random	0.313	**0.229**	0.307
Ti_4w_5R	Sphere	3.103	2.236	**1.078**
Ti_4w_5R	Overall	1.928	0.946	**0.706**

* MBS-OPT is empirically optimized by this dataset.

Table 2. Evaluation of VolRAFT to Bruns method [7] using the default vanilla (MBS-VAN) and the optimal setting (MBS-OPT), based on average End-Point-Error (EPE). The best is highlighted.

and the fields estimated by MBS-VAN, MBS-OPT and VolRAFT using the synthetic datasets Mg-5Gd_4w_103L and the Star field. The screw volume, which was not evaluated, is shown in black in the image. Results show that the displacement predicted by VolRAFT generally resembles the displacement as shown within the ground truth. By contrast, the classical iterative method would require further fine-tuning of the smoothness term and prior knowledge to handle continuous fields such as this Star field. After empirical optimization for Bruns's method by the objective EPE value and its qualitative accuracy, MBS-OPT only shows slight improvement to its vanilla setting MBS-VAN.

However, VolRAFT can obtain block-based artifacts near the edge of homogeneous region (see the f_h field, Fig. 3 middle-row last-column). This is due to the fact that VolRAFT utilizes patch-based inference methods, which can lead to a severe jump of values near the edge of patches. We have employed Gaussian weights to blend and suppress these artifacts (as shown in Sec. 3), but the nature of this patch-based inference cannot be completely eliminated. One common practise is to include regularization terms such as total variation utilized by Bruns's method. However, the use of regularization terms introduces the need for manual tuning thereof by the user to the application of the deep learning model. As the aim of this work is to present a methodology for DVC without user input, we exclude the regularization from the scope in this paper.

To examine the quality in a realistic measurement environment, we test VolRAFT by inference to the experimental measured volume pairs based on the model parameters trained by synthetic datasets. Fig. 4 shows the displacement fields obtained by MBS-VAN, MBS-OPT and VolRAFT using real measured volumes of PEEK_4w_5L sample at its first and second scan as the reference and deformed volume, respectively. Similarly, Fig. 5 shows the displacement fields based on MBS-VAN, MBS-OPT and VolRAFT of a
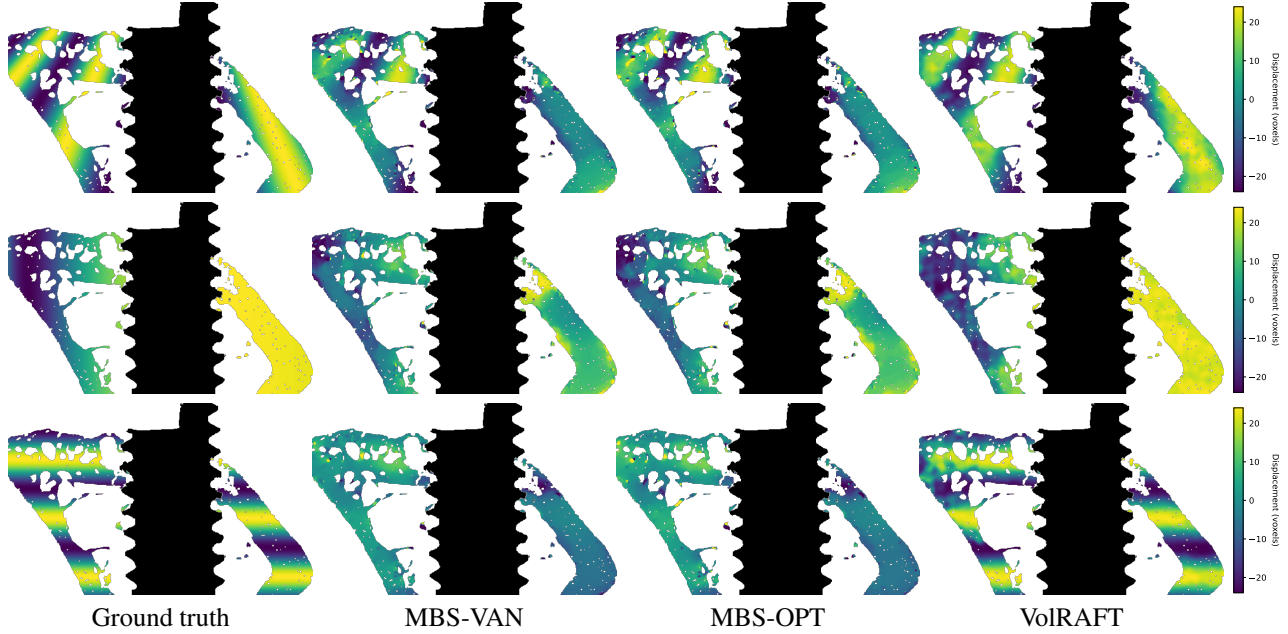
Figure 3. Comparison of the ground truth (first column) and displacement fields estimated by: (second column) MBS-VAN [7]; (third column) MBS-OPT; (last column) the proposed VolRAFT approach. Displacement fields (in voxels) are shown in the direction of: width f_w (top-row), height f_h (middle-row) and depth f_d (bottom-row), respectively. Cross sections (slices) of displacement fields are shown here.
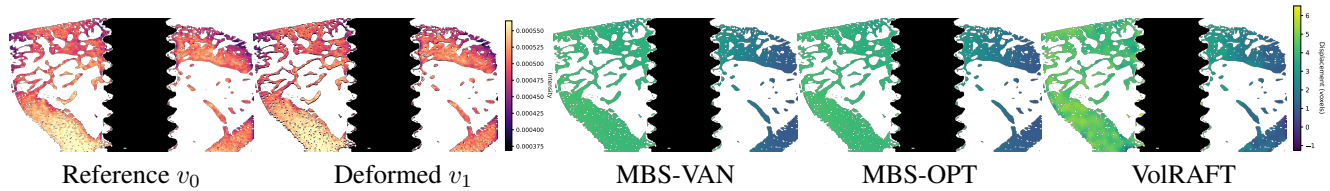

Figure 4. Evaluation of the measured volumes of PEEK_4w_5L, based on the reference volume (first) and the real deformed volume (second). Displacement fields are estimated by (from left to right): MBS-VAN [7] (third); MBS-OPT (fourth); the proposed VolRAFT approach (fifth, the last one at the right). Cross sections (slices) of volumes and displacement fields are shown here.

Ti_4w_5R sample at its first and last scan as the reference (Fig. 5a) and deformed (Fig. 5b) volume. We observe that VolRAFT essentially estimates the structure of displacement fields correctly, if we consider the Bruns's method as the optimal solution of DVC for this material class. However, as it is extremely difficult to measure the ground truth of a realistic displacement field in a SRμCT experimental setup, we can only evaluate these results in a qualitative manner. The generalization capability of VolRAFT is particularly impressive, as both samples are completely unknown in the training set.

Within the Ti_4w_5R sample we can observe a fracture (magnified in the circle box of Fig. 5) in the deformed volume that extends from the screw into the cortical bone. This presents a particular challenge in DVC, as the volumetric structure of this fracture (see the magnified region Fig. 5b) is an unknown pattern not present in the reference volume. The fracture results in a discontinuous and sharp change

such as the edge observed in the displacement field (magnified in the circle box of Fig. 5). In the proposed VolRAFT approach, we did not particularly address this scenario by generating any corresponding synthetic field. Surprisingly, the estimated field by VolRAFT (bottom-right, Fig. 5) can roughly estimate this structure nonetheless. An artifact is visible within the flow field, which is likely caused by the patch blending during the inference. This capability of VolRAFT is potentially due to the training on the synthetic Sphere field, as it includes scenarios of positive divergence (*i.e.* a point source of 3D vector fields). The performance of VolRAFT for monitoring crack formation needs to be further studied in the future.

To further test the generalizability of VolRAFT, future studies should include displacements based on individual motions of multiple objects and displacements of objects with little inside texture, as can be the case in mechanical testing of powder and particle samples and tensile testing of

(a) Reference volume v_0 (b) Deformed volume v_1

MBS-VAN   MBS-OPT   VolRAFT

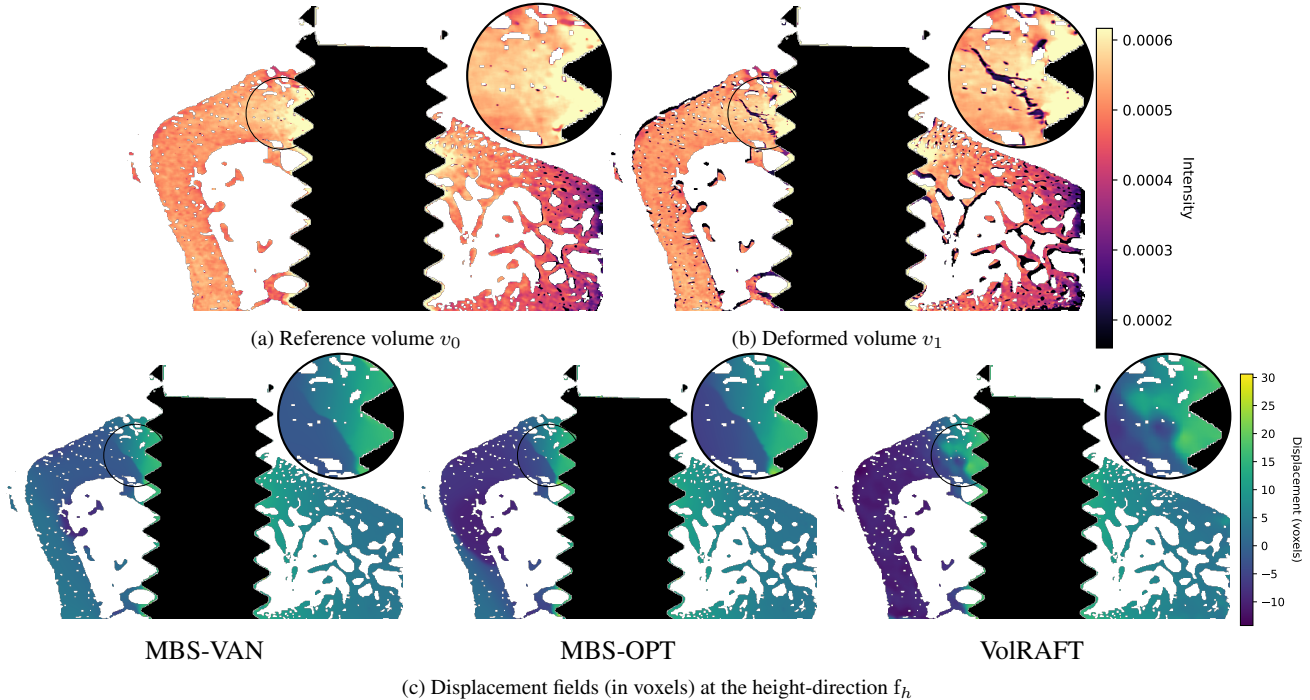(c) Displacement fields (in voxels) at the height-direction f_h

Figure 5. Evaluation of the measured volumes of `Ti_4w_5R`, based on the reference volume (Fig. 5a) and the real deformed volume (Fig. 5b). Displacement fields (Fig. 5c) are estimated by: (left column) MBS-VAN [7]; (middle column) MBS-OPT; (right column) the proposed VolRAFT approach. Cross sections (slices) of volumes and displacement fields are shown here. The fracture is magnified in the circle area at volumes and displacement fields.

metallic samples, respectively.

6. Conclusion

In this paper, we propose the VolRAFT approach that extends the optical flow network RAFT from 2D images to 3D volumes, without further fine-tuning or optimizing its hyperparameters. The experiments demonstrates that this approach generally performs better using synthetic displacement fields with strong and significant divergence and curl in comparison to cutting-edge iterative methods. As the realistic measurement datasets show, the generalizability is particularly impressive, as it can still resemble the overall structure of the field even if the testing set is completely unknown in the training set of VolRAFT model.

This work provides insight into application of state-of-the-art computer vision approaches to address classical challenges in materials science research, such as the DVC analysis shown in this paper. In the future, this approach can be improved by its inference method, and it can be potentially applied to other materials and experimental data.

Ethics statement

Experiments involved with animals were approved by the ethical committee at the Malmö/Lund regional board for an-

imal research, Swedish Board of Agriculture (approval no. DNR M 188-15). Authors declare no competing interests.

References

[1] Simon Baker, Daniel Scharstein, James P Lewis, Stefan Roth, Michael J Black, and Richard Szeliski. A database and evaluation methodology for optical flow. *IJCV*, 92:1–31, 2011. 6

[2] E Bar-Kochba, J Toyjanova, E Andrews, K-S Kim, and Christian Franck. A fast iterative digital volume correlation algorithm for large deformations. *Experimental Mechanics*, 55:261–274, 2015. 1, 2

[3] Brian K Bay. Methods and applications of digital volume correlation. *The Journal of Strain Analysis for Engineering Design*, 43(8):745–760, 2008. 2

[4] Michel Bornert, Fabrice Brémand, Pascal Doumalin, J-C Dupré, Marina Fazzini, Michel Grediac, François Hild, Sebastien Mistou, Jérôme Molimard, J-J Orteu, et al. Assessment of digital image correlation measurement errors: methodology and results. *Experimental mechanics*, 49:353–370, 2009. 5

[5] Seyfeddine Boukhtache, Kamel Abdelouahab, François Berry, Benoît Blaysat, Michel Grediac, and Frédéric Sur. When deep learning meets digital image correlation. *Optics and Lasers in Engineering*, 136:106308, 2021. 5

[6] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In *Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part IV 8*, pages 25–36. Springer, 2004. 2

[7] Stefan Bruns, Diana Krüger, Silvia Galli, DC Florian Wieland, Jörg U Hammel, Felix Beckmann, Ann Wennerberg, Regine Willumeit-Römer, Berit Zeller-Plumhoff, and Julian Moosmann. On the material dependency of peri-implant morphology and stability in healing bone. *Bioactive Materials*, 28:155–166, 2023. 1, 2, 4, 5, 6, 7, 8

[8] Ante Buljac, Clément Jailin, Arturo Mendoza, Jan Neggers, Thibault Taillandier-Thomas, Amine Bouterf, Benjamin Smaniotto, François Hild, and Stéphane Roux. Digital volume correlation: review of progress and challenges. *Experimental Mechanics*, 58:661–708, 2018. 1, 2

[9] Enrico Dall'Ara and Gianluca Tozzi. Digital volume correlation for the characterization of musculoskeletal tissues: Current challenges and future developments. *Frontiers in Bioengineering and Biotechnology*, page 1867, 2022. 1, 2

[10] CM Disney, J Mo, A Eckersley, AJ Bodey, JA Hoyland, MJ Sherratt, AA Pitsillides, PD Lee, and BK Bay. Regional variations in discrete collagen fibre mechanics within intact intervertebral disc resolved using synchrotron computed tomography and digital volume correlation. *Acta Biomaterialia*, 138:361–374, 2022. 2

[11] Tomy dos Santos Rolo, Alexey Ershov, Thomas van de Kamp, and Tilo Baumbach. In vivo x-ray cine-tomography for tracking morphological dynamics. *Proceedings of the National Academy of Sciences*, 111(11):3921–3926, 2014. 1, 2

[12] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. In *ICCV*, pages 2758–2766, 2015. 2

[13] Xiaocen Duan and Jianyong Huang. Deep learning-based digital volume correlation. *Extreme Mechanics Letters*, 53: 101710, 2022. 2, 5

[14] Denis Fortun, Patrick Bouthemy, and Charles Kervrann. Optical flow modeling and computation: A survey. *Computer Vision and Image Understanding*, 134:1–21, 2015. 2, 6

[15] Simon Hermann and René Werner. High accuracy optical flow for 3d medical image registration using the census cost function. In *Image and Video Technology: 6th Pacific-Rim Symposium, PSIVT 2013, Guanajuato, Mexico, October 28-November 1, 2013. Proceedings 6*, pages 23–35. Springer, 2014. 1, 2

[16] John Holmes, Silvano Sommacal, Raj Das, Zbigniew Stachurski, and Paul Compston. Digital image and volume correlation for deformation and damage characterisation of fibre-reinforced composites: A review. *Composite Structures*, page 116994, 2023. 1

[17] Zhaoyang Huang, Xiaoyu Shi, Chao Zhang, Qiang Wang, Ka Chun Cheung, Hongwei Qin, Jifeng Dai, and Hongsheng Li. Flowformer: A transformer architecture for optical flow. In *ECCV*, pages 668–685. Springer, 2022. 3

[18] Amira I Hussein, Paul E Barbone, and Elise F Morgan. Digital volume correlation for study of the mechanics of whole bones. *Procedia Iutam*, 4:116–125, 2012. 1, 2

[19] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. Flownet 2.0: Evolution of optical flow estimation with deep networks. In *CVPR*, pages 2462–2470, 2017. 2

[20] Di Jia, Kai Wang, ShunLi Luo, TianYu Liu, and Ying Liu. Braft: Recurrent all-pairs field transforms for optical flow based on correlation blocks. *IEEE Sign. Process. Letters*, 28:1575–1579, 2021. 3

[21] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2018. 5

[22] Kamel Madi, Katherine A Staines, Brian K Bay, Behzad Javaheri, Hua Geng, Andrew J Bodey, Sarah Cartmell, Andrew A Pitsillides, and Peter D Lee. In situ characterization of nanoscale strains in loaded whole joints via synchrotron x-ray tomography. *Nature biomedical engineering*, 4(3):343–354, 2020. 2

[23] Tessa Nogatz, Claudia Redenbach, and Katja Schladitz. 3d optical flow for large ct data of materials microstructures. *Strain*, 58(3):e12412, 2022. 1, 2

[24] Julian Sartori, Sebastian Köhring, Stefan Bruns, Julian Moosmann, and Jörg U Hammel. Gaining insight into the deformation of achilles tendon entheses in mice. *Advanced Engineering Materials*, 23(11):2100085, 2021. 2

[25] Adrian A Schiefler, Stefan Bruns, Dirk Müter, Kentaro Uesugi, Henning Osholm Sørensen, and Dominique J Tobler. Retention of sulfidated nzvi (s-nzvi) in porous media visualized by x-ray μ-ct–the relevance of pore space geometry. *Environmental Science: Nano*, 9(9):3439–3455, 2022.

[26] Sebastian Schmelzle, Stefan Bruns, Felix Beckmann, Julian Moosmann, and Silke Lautner. Using in situ synchrotron-radiation-based microtomography to investigate 3d structure-dependent material properties of tension wood. *Advanced Engineering Materials*, 23(11):2100235, 2021. 2

[27] Sabrina Chin-yun Shen, Marta Peña Fernández, Gianluca Tozzi, and Markus J Buehler. Deep learning approach to

assess damage mechanics of bone tissue. *Journal of the Mechanical Behavior of Biomedical Materials*, 123:104761, 2021. 2

[28] Xiaoyu Shi, Zhaoyang Huang, Dasong Li, Manyuan Zhang, Ka Chun Cheung, Simon See, Hongwei Qin, Jifeng Dai, and Hongsheng Li. Flowformer++: Masked cost volume autoencoding for pretraining optical flow estimation. In *CVPR*, pages 1599–1610, 2023. 3

[29] Xiuchao Sui, Shaohua Li, Xue Geng, Yan Wu, Xinxing Xu, Yong Liu, Rick Goh, and Hongyuan Zhu. Craft: Cross-attentional flow transformer for robust optical flow. In *CVPR*, pages 17602–17611, 2022. 2, 3

[30] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *CVPR*, pages 8934–8943, 2018. 2

[31] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *ECCV*, pages 402–419. Springer, 2020. 1, 2, 3, 4

[32] Domonkos Tolnai, Sarkis Gavras, Fabian Wilde, Jörg U Hammel, and Stefan Bruns. In situ synchrotron tomography of the solidification of an elektron 21 mg alloy. *Advanced Engineering Materials*, 23(11):2100383, 2021. 2

[33] Zhigang Tu, Wei Xie, Dejun Zhang, Ronald Poppe, Remco C Veltkamp, Baoxin Li, and Junsong Yuan. A survey of variational and cnn-based optical flow techniques. *Signal Processing: Image Communication*, 72:9–24, 2019. 2

[34] Erika Tudisco, Edward Andò, Rémi Cailletaud, and Stephen A Hall. Tomowarp2: A local digital volume correlation code. *SoftwareX*, 6:267–270, 2017. 2

[35] Guowen Wang, Laibin Zhang, and Xuefeng Yao. Strainnet-3d: Real-time and robust 3-dimensional speckle image correlation using deep learning. *Optics and Lasers in Engineering*, 158:107184, 2022. 2

[36] Mingliang Zhai, Xuezhi Xiang, Ning Lv, and Xiangdong Kong. Optical flow and scene flow estimation: A survey. *Pattern Recognition*, 114:107861, 2021. 1, 2