Digital image correlations and deep learning: new perspectives for experimental dynamics

N. N. Balaji [1], **M. R. W. Brake** [1], **C. M. Jermaine** [2]
[1] Rice University, Department of Mechanical Engineering,
Houston, Texas, USA

[2] Rice University, Department of Computer Science,
Houston, Texas, USA

Abstract
Digital Image Correlation (DIC) is becoming increasingly popular since it presents a full field non-contact experimental approach that is applicable in a variety of contexts. Since it necessitates the use of high-speed cameras (for experimental vibrations), one needs to contend with the trade-off between spatial resolution, temporal sampling, and maximum test duration. The present contribution presents an investigation of the employment of deep learning for video-based experimental dynamics. A class of deep learning strategies are employed to determine if an appropriately trained model can substitute DIC completely in an experimental setting. Recent literature pertaining to quasi-static experiments have documented encouraging results where such models have been shown to even outperform DIC. Networks using 2D convolutional ResNets (applied on pairs of frames) are considered. Random vibration FRF testing conducted on beam-like structures is used as an experimental benchmark to bring out the pros and cons of this approach.

1 Introduction

Digital Image Correlation, or DIC in short, has become a well-established full field experimental measurement approach, especially in the structural mechanics community [1]. DIC has made it possible to use optical measurements, i.e., frames from a video, to generate displacement and strain fields with comparable accuracy as traditional contact sensing approaches involving strain gauges, PZTs, etc. DIC works by parameterizing frame-to-frame spatial correlations of the images (taken as a set of subsets) by a spatially piece-wise displacement field and optimizing these parameters against the measured frames taken in sequence. The approach has seen applicability in a vast variety of applications (see [2] for a recent review) including, more recently, experimental modal analysis [3] (EMA).

One challenge with employing DIC for EMA lies in the need to synchronize the excitation (through an impact hammer or an electrodynamic shaker) and the acquisition of the frames of video. Although this can be managed using programmable acquisition systems, there exists a second, more acute issue. Most high-speed cameras are limited by the amount of onboard memory, restricting the total number of pixels that can be stored at any time. This places rather stringent constraints on the number of averages one can do, for instance, for frames at a fixed resolution. It goes without saying that the framerate limitations of the camera offer further constraints. Apart from these, owing to the large quantity of data that needs to be processed for extracting information from such tests, the applicability of DIC as a real-time diagnostics/sensing tool is fundamentally limited.

More recently, spurred by the successes in the deep learning community, studies have started appearing in the literature applying deep learning to either assist or completely substitute DIC [4, 5, 6, 7]. While the earlier studies [4, 5] focused on training deep neural networks to replicate the results of classical and obtained results of varying accuracy and applicability, more recent ones [6, 7] have focused on employing synthetic displacement fields to train the network. Ref. [7], for instance, have demonstrated that using such

an approach, deep neural networks can provide comparable or even higher accuracy in the displacement and strain field predictions in comparison to commonly available DIC software (VIC-2D, ncorr, Sandia's DICe, etc.). We will use this as the starting point for the current study.

The primary focus of the present study lies in the applicability of the deep learning network developed in [7] for a dynamic test case involving data from EMA tests of an Aluminum beam undergoing random excitation by an electrodynamic shaker (see section 2.1 for details). The local acceleration predictions will be compared against data from a traditional accelerometer in both the time as well as frequency domains (FRF estimates). The rest of the paper is organized into Methodology (in section 2), describing the deep learning framework under consideration and the experimental setup; Results (in section 3) presenting the results of the application of the deep learning approaches; and finally concludes with key discussions and observations (in section 6).

2 Methodology

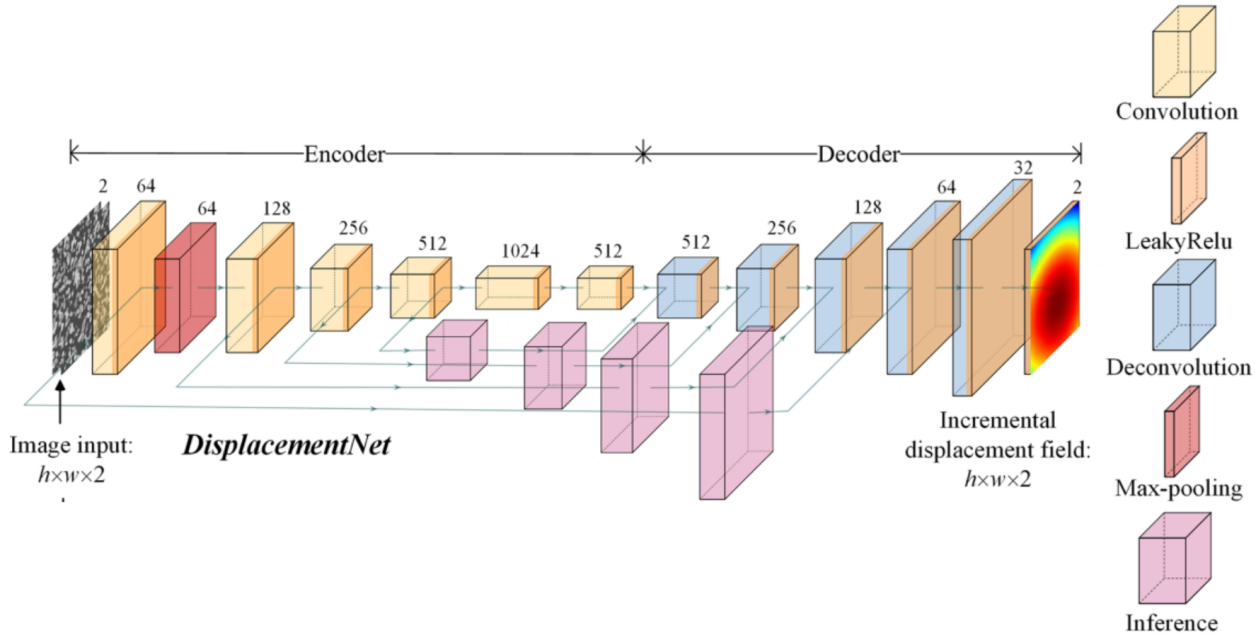


Figure 1: DeepDIC DispNet Architechture (figure from Ref. [7])

In [7], an encoder-decoder architecture inspired by the so-called FlowNet [8], which is a very popular deep neural network architecture for optical flow, was train to predict displacement and strain fields from speckled image inputs. The network, shown in fig. 1, takes pairs of images (reference and deformed) as input, and outputs horizontal and vertical displacement fields. Although the same has been done for strain predictions too, we will restrict ourselves to the DispNet for the present study.

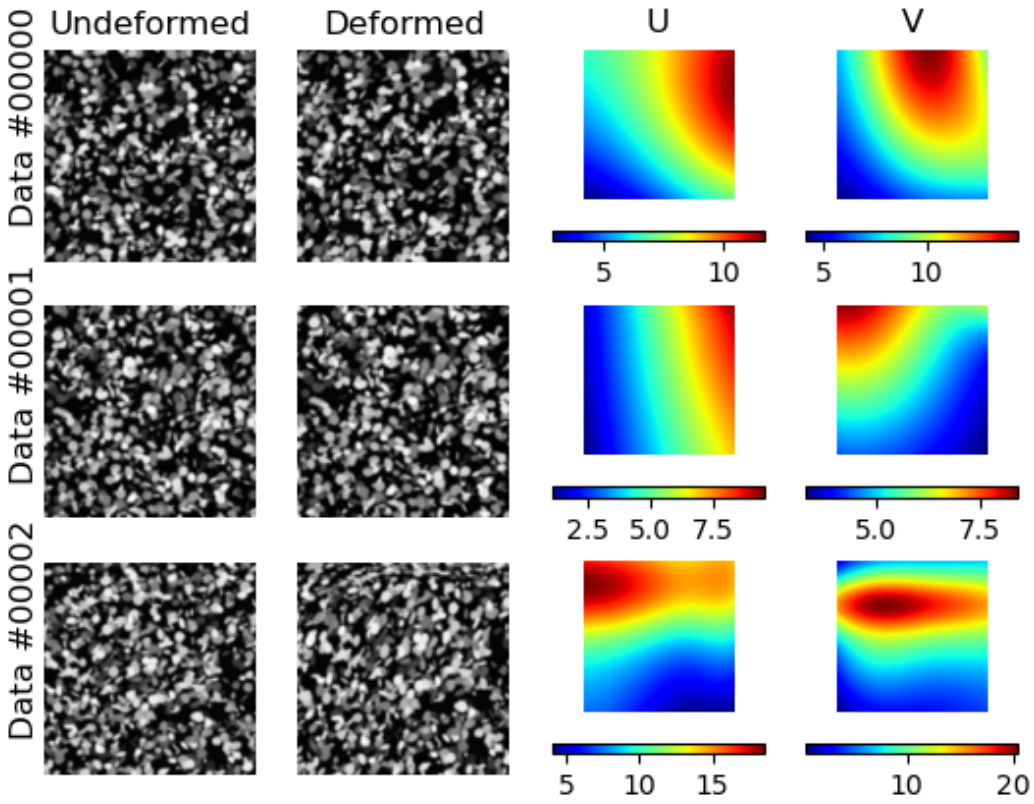
The network is trained using synthetically datasets generated in a manner similar to that described in [7], but with parametric modifications that reflect the displacement (amplitude) regimes expected from the present tests better. 40,000 128×128 pixel synthetic image-pairs are generated for the current study with speckle patterns generated by arbitrarily oriented ellipses on a plane. Figure 2 shows the displacement distributions as well as three sample data pairs and their displacement fields (shown in pixel units). This dataset is used to finetune the model published by the authors [1].

2.1 Experimental setup

An aluminum beam of dimensions $354mm \times 25mm \times 6mm$ is used for the studies in this paper. A photograph of the experimental setup is shown in Figure 5. The electrodynamic shaker is excited using band-limited multi-sine signals through National Instruments serial cDAQ hardware, programmed using LabView. A Phantom high speed camera is used for the acquisition. In order to synchronize the frames measured by the camera and the vibration acquisition, trigger pulses are sent to an auxiliary trigger channel of the camera to indicate the start of the experiment, after which both the cDAQ as well as the camera measure a fixed number of samples/frames. All experiments are conducted with a sampling of 10.24 kHz, unless otherwise mentioned.

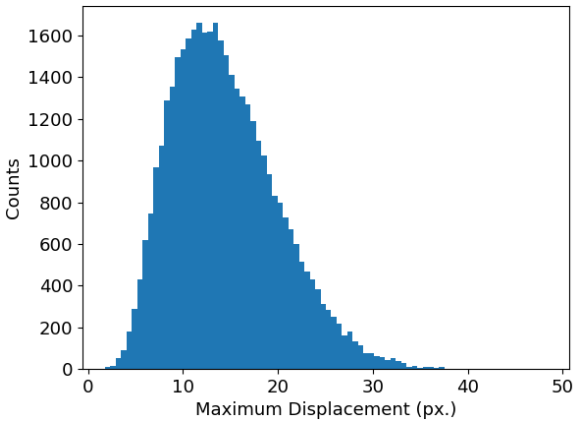
Two types of experiments are conducted, with the camera focused on different regions of the beam, as shown in fig. 4. As can be seen in the figure, different speckling techniques are used for both of these to ensure that there are sufficient features in the images for the model to track.

[1] see `https://github.com/RuYangNU/Deep-Dic-deep-learning-based-digital-image-correlation`

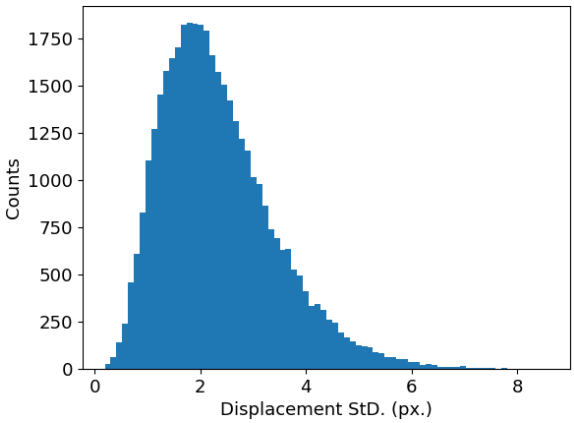


(a)

Figure 2: (Contd. 1/2) Synthetic DataSet generated for the study: (a) sample image pairs along with their displacement fields (in pixel units); (b-c) histograms of the maximum and standard deviations of the displacement fields across the dataset.



(b)



(c)

Figure 2: (Contd. 2/2) Synthetic DataSet generated for the study: (a) sample image pairs along with their displacement fields (in pixel units); (b-c) histograms of the maximum and standard deviations of the displacement fields across the dataset.

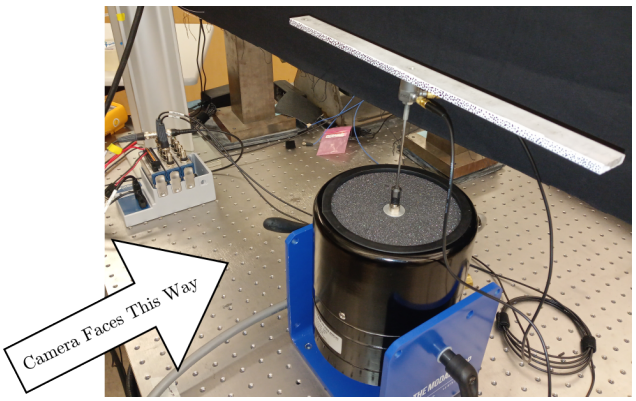
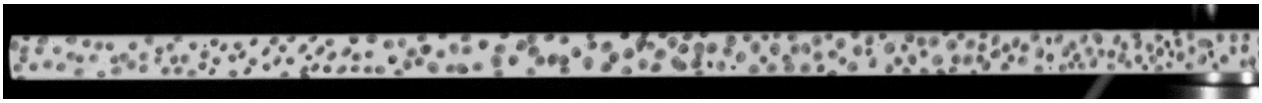
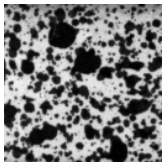


Figure 3: Experimental setup for the Aluminum beam



(a)



(b)

Figure 4: The reference frame for the two tests performed: (a) Test 1, with the frame focusing on one half of the beam; and (b) Test 2, with the frame focusing on the left end of the beam.

3 Results

Figure 5 presents the training and validation losses. The model is implemented in PyTorch [9] and an ADAM optimizer (as implemented in PyTorch) is used for the training. The learning rate was fixed by conducting cross-validation with single epochs training steps conducted at regular intervals. For this case, a learning rate

of 3×10^{-3} was fixed for the first 50 epochs and then reduced to 2×10^{-4} for the last 100 epochs. In fig. 5 the MSE loss has as units pixels-squared, and the validation error is around $10^{-1} px^2$, while the training error is even smaller. This demonstrates that the model is now in a regime where it has low generalizability, i.e., it is being limited by the dataset. This indicates that the size of the dataset is not sufficient and has to be increased. The authors are presently investigating an "infinite dataset approach" to check if the validation errors improve. SANDIA's open source Digital Image Correlation engine is employed for all the DIC studies unless otherwise mentioned.

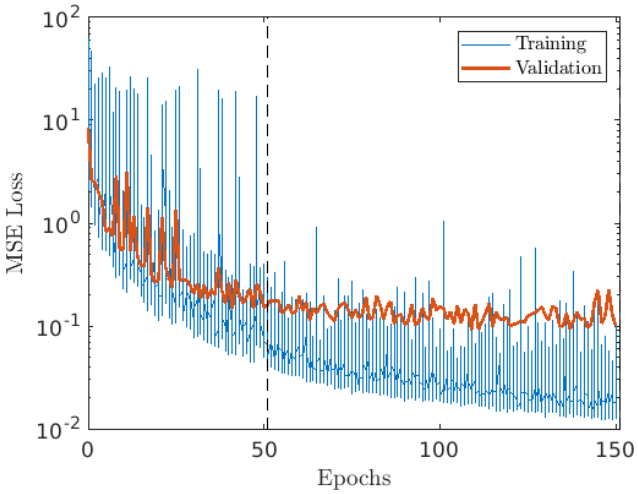


Figure 5: Training and Validation errors by epoch during finetuning. A learning rate of 3×10^{-3} was first used, and this was reduced to 2×10^{-4} after 50 epochs.

4 EMA results

Figure 6 shows the results of the first test conducted on the aluminum beam. The frame size is chosen to be 94×1280 pixels. A total of 14 blocks, each with 6400 samples, are recorded, with the first two blocks representing zero excitation (for noise estimation), and the last 12 blocks representing three different repeats of random excitation, each conducted for four cycles (to remove transients). The Frequency Response Functions plotted in fig. 6(b) show a good match between the DIC predictions and the accelerometer data for the low frequency regime (until around 1000Hz), but the uncertainty in the DIC really starts accumulating at higher frequencies than this.

5 Performance of the finetuned Deep-DIC for EMA

Figure 7 shows time-domain comparisons of the deep-learning model's predictions against that of the classical DIC at approximately the same point in space. It can be seen that the DDIC approach predicts the time-series in a manner that is nearly similar to the classical DIC, although it was trained fully independent of DIC. This independence is also brought out by the fact that the DDIC predictions seem to have some higher harmonics present in the response, which DIC seems to lack.

Figure 8 compares the frequency responses estimated with the data from the DIC, DDIC, and the accelerometer measurements along with their uncertainties. It can be seen that the uncertainties of the DIC and DDIC approaches are approximately similar, but the FRF predictions from the DDIC approach are only good around the peak, showing that the DDIC approach performs rather poorly in the frequency domain.

Similarly, fig. 9 presents the time and frequency domain results for test 2. In this case the DIC and DDIC predictions match near exactly. This could be due to the fact that the speckle pattern opted for this case is more similar to the training data than in test 1. In this case no accelerometer was placed near the end of

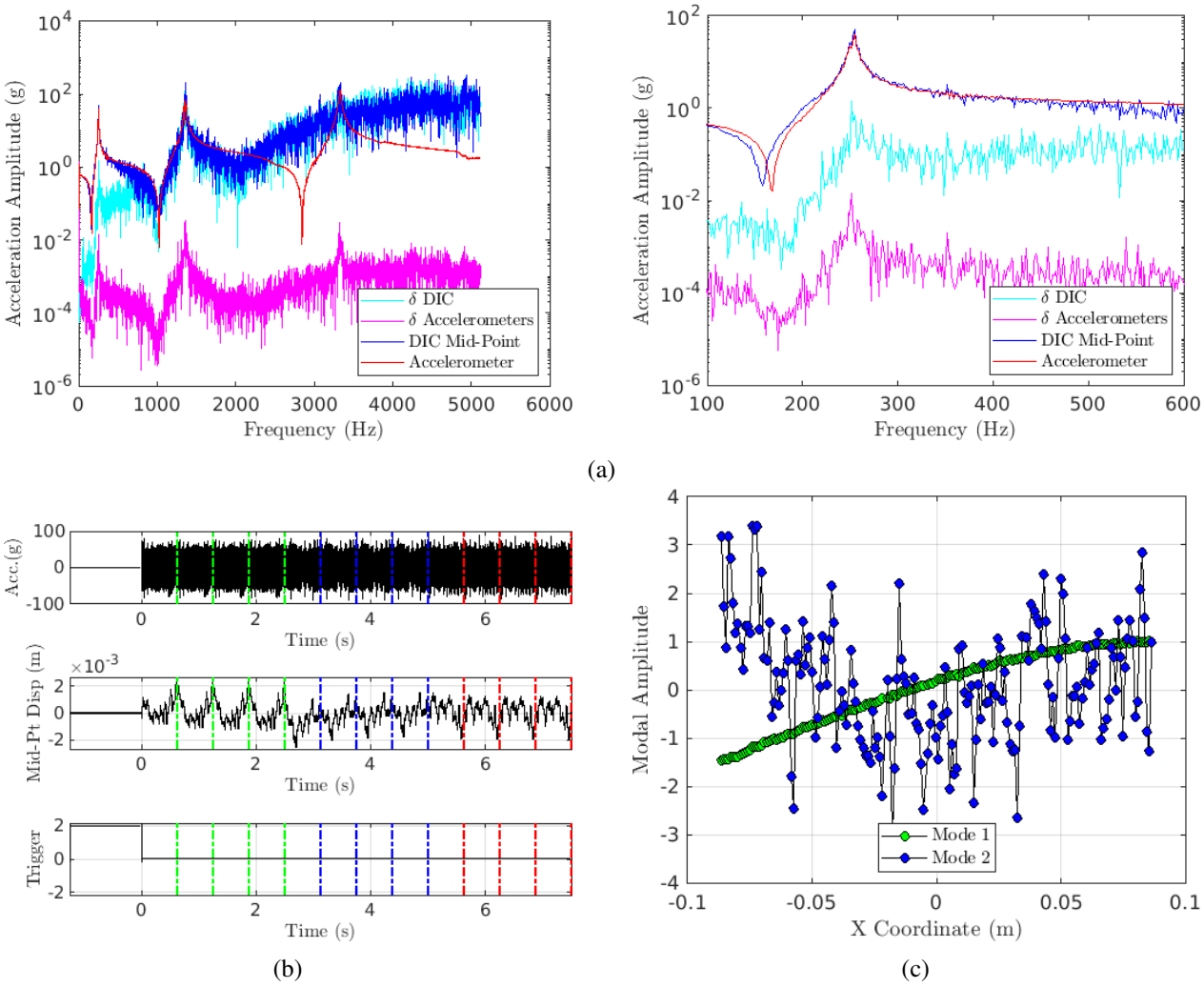


Figure 6: Results from the EMA carried out using DIC for Test 1: (a) frequency response estimates from a vibration accelerometer compared against DIC. Also plotted are the uncertainties (standard deviations from averaging) for each case; (b) time-domain view of acceleration, displacement and trigger. The different blocks are separated by vertical lines with the colors denoting the repeats; and (c) the first (green) and second (blue) mode shapes estimated from DIC against.

the beam so only the DIC results are shown (in pixel units). Once again it can be observed that there is significant uncertainty present for frequencies more than around 300Hz.

Table 1: Total computational processing time taken for DIC and DDIC for the two tests

	Frame Size	Num. Frames	DIC	DDIC	Speed-up
Test 1	96×1280	89,600	22,750 s	12,527 s	1.8x
Test 2	128×128	652,800	247,500 s	17,033 s	14.53x

Lastly, the computational efforts involved in employing DIC and DDIC are compared in table 1 along with the net speedup that DDIC provides. It can be seen that in both the cases the DDIC approach offers a good speed-up in comparison to DIC. However, this seems to be critically dependent on the frame size. The scaling of the convolutional neural network in comparison with the scaling of the DIC implementation with frame size has to be investigated further in order to determine in which cases the use of DDIC provides a considerable advantage over DIC.

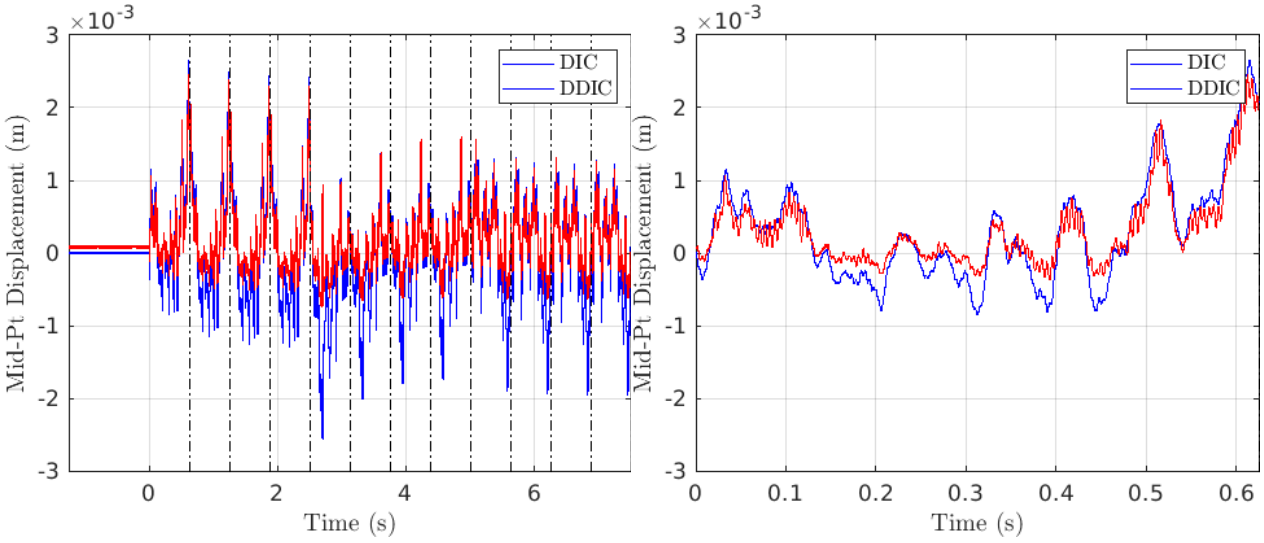


Figure 7: Time domain comparisons of the DIC and Deep-DIC (DDIC) predictions for test1

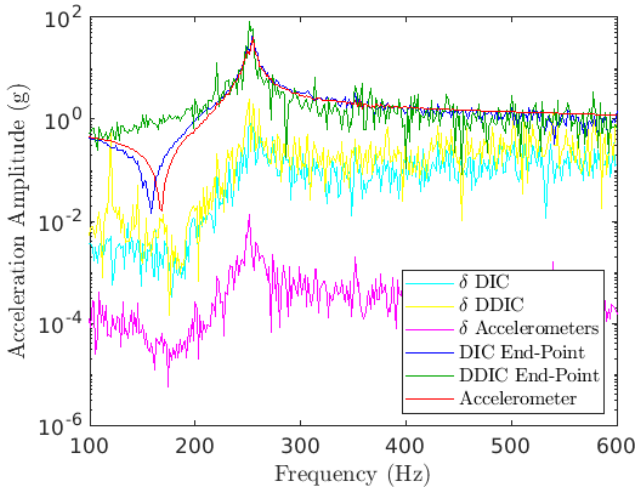


Figure 8: FRF comparison of the estimations carried out using DIC, DDIC, and accelerometer measurements along with uncertainty estimates.

6 Conclusions

A Deep Convolutional Neural Network was finetuned and applied for experimental modal analysis. The first conclusion is that the Deep-DIC (DDIC) approach seems to be, at worst, a very computationally efficient alternative to DIC. This is brought out by the very promising accuracy of DDIC while offering up to around 15x speed-ups in computation times. However, the performance of both the DIC & DDIC for EMA is still quite unsatisfactory and is the subject of ongoing research.

One aspect that requires improvement is the fine-tuning of the network. A clear training-validation divergence was observed during this study, indicating that an "infinite dataset" formulation could offer better results. Another aspect is that the neural network used here merely compares a single reference frame with multiple "deformed" frames, i.e., the DDIC is applied pair-wise, with no dynamics built in. Alternative architectures that can capture the spatio-temporal correlations (like 3D CNNs) can be explored in the future for improvements.

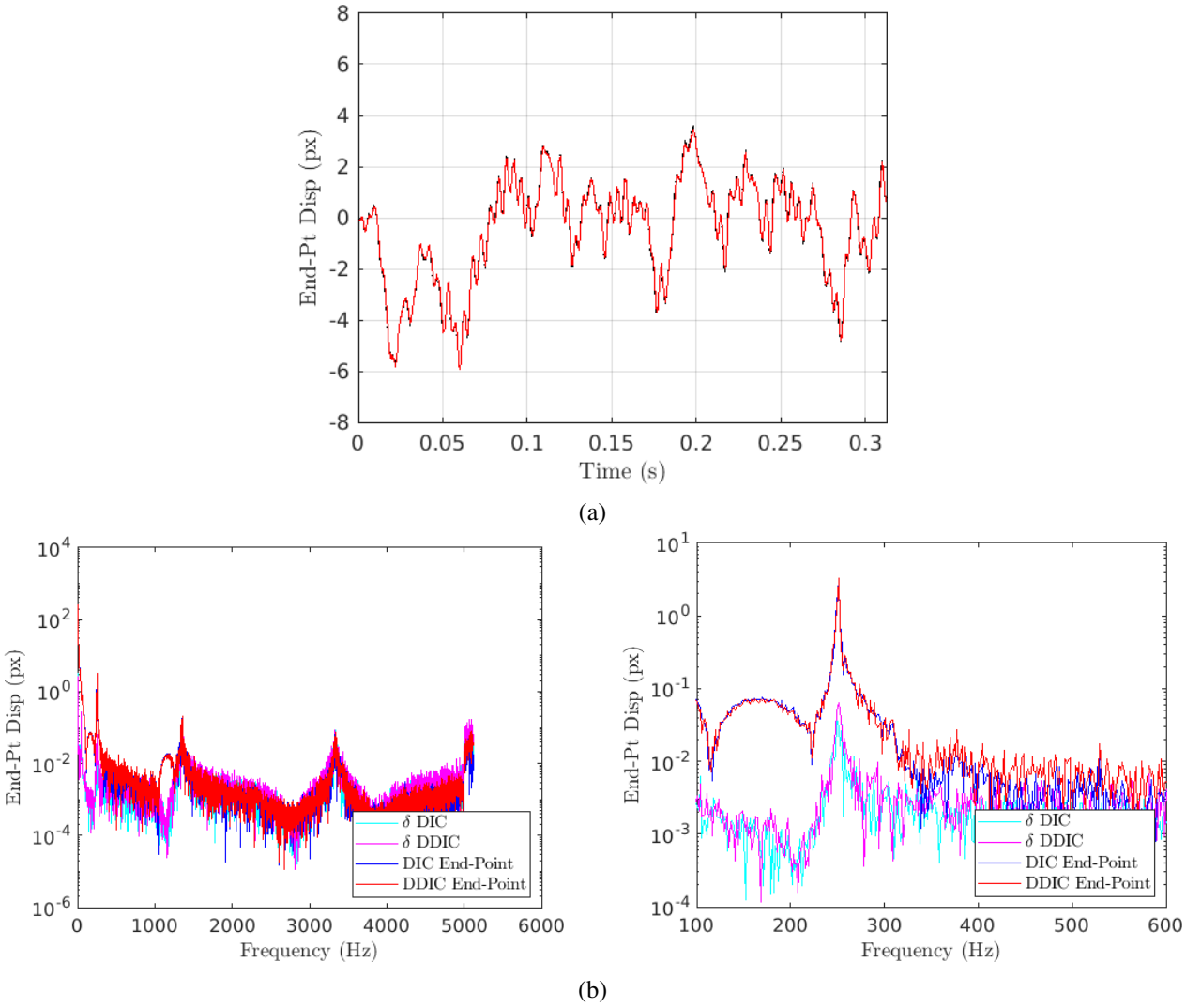


Figure 9: (a) Time domain and (b) Frequency responses from test 2 comparing the results from DIC and the Deep-DIC implementation.

References

[1] P. Reu, "Introduction to Digital Image Correlation: Best Practices and Applications," *Experimental Techniques*, vol. 36, no. 1, pp. 3–4, Jan. 2012.

[2] B. Pan, "Digital image correlation for surface deformation measurement: Historical developments, recent advances and future goals," *Measurement Science and Technology*, vol. 29, no. 8, p. 082001, Aug. 2018.

[3] E. Di Lorenzo, D. Mastrodicasa, L. Wittevrongel, P. Lava, and B. Peeters, "Full-Field Modal Analysis by Using Digital Image Correlation Technique," in *Rotating Machinery, Optical Methods & Scanning LDV Methods, Volume 6*, D. Di Maio and J. Baqersad, Eds. Cham: Springer International Publishing, 2020, pp. 119–130.

[4] H.-G. Min, H.-I. On, D.-J. Kang, and J.-H. Park, "Strain measurement during tensile testing using deep learning-based digital image correlation," *Measurement Science and Technology*, vol. 31, no. 1, p. 015014, Oct. 2019.

[5] A. Rezaie, R. Achanta, M. Godio, and K. Beyer, "Comparison of crack segmentation using digital image correlation measurements and deep learning," *Construction and Building Materials*, vol. 261, p. 120474, Nov. 2020.

[6] S. Boukhtache, K. Abdelouahab, F. Berry, B. Blaysat, M. Grédiac, and F. Sur, "When Deep Learning Meets Digital Image Correlation," *Optics and Lasers in Engineering*, vol. 136, p. 106308, Jan. 2021.

[7] R. Yang, Y. Li, D. Zeng, and P. Guo, "Deep DIC: Deep learning-based digital image correlation for end-to-end displacement and strain measurement," *Journal of Materials Processing Technology*, vol. 302, p. 117474, Apr. 2022.

[8] P. Fischer, A. Dosovitskiy, E. Ilg, P. Häusser, C. Hazırbaş, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox, "FlowNet: Learning Optical Flow with Convolutional Networks," *arXiv:1504.06852 [cs]*, May 2015.

[9] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf