Time-varying implicit neural representations for unsupervised speckle denoising in dynamic scenes

Matthew R. Ziemann[*,a,b], Rushil R. Joshi[*,b], and Christopher A. Metzler[b]

[a]DEVCOM Army Research Laboratory, Adelphi, MD, USA
[b]University of Maryland, College Park, MD, USA
[*]Contributed equally

ABSTRACT

Speckle noise is inherent to coherent imaging systems such as synthetic aperture radar (SAR), optical coherence tomography (OCT), and ultrasound imaging. However, its multiplicative nature makes it especially challenging to remove. Today the most effective speckle denoising methods average multiple identically distributed measurements—however, these approaches fail to reconstruct dynamic scenes. In this work we leverage implicit neural representations (INRs) to perform unsupervised speckle denoising of *time-varying* sequences. We optimize a maximum likelihood-based loss function to produce high-fidelity, speckle-free reconstructions. Our approach significantly outperforms existing techniques, achieving up to a 4 dB improvement in peak signal-to-noise ratio (PSNR) for dynamic scenes with simulated speckle.

Keywords: speckle, denoising, implicit neural representations, coherent imaging, computational imaging, deep learning

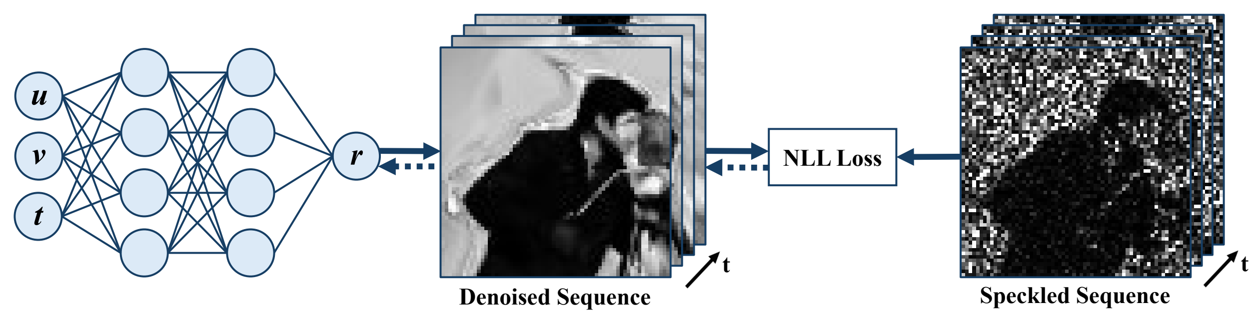

Figure 1. **Denoising Dynamic Scenes.** Illustration of our denoising framework, featuring a multi-layer perceptron (MLP)-based INR optimizing our negative log likelihood loss given a time-varying sequence of speckled measurements.

1. INTRODUCTION

Speckle noise is inherent to coherent imaging systems such as synthetic aperture radar (SAR), optical coherence tomography (OCT), and ultrasound imaging, where coherent light or sound waves interact with rough surfaces or heterogeneous media.[1] In these domains, speckle noise manifests as granular interference patterns that spatially vary in intensity depending on the reflectivity of the scene. Due to its multiplicative nature, speckle can be more challenging than more traditional noise models like Gaussian or Poisson, and this challenge increases with smaller aperture sizes that increase the spatial correlation of the noise.[2]

In dynamic scenes, where either the sensor or the target is in motion, speckle noise becomes even more problematic. Traditional speckle denoising techniques often rely on static or quasi-static assumptions and typically

Further author information:
M.R.Z.: matthew.r.ziemann.civ@army.mil; R.R.J.: rjoshi23@umd.edu; C.A.M.: metzler@umd.edu

require multiple identical measurements of a static scene.[2] We address this gap by proposing a novel approach leveraging implicit neural representations to denoise speckle from time-varying sequences, as depicted in Figure 1.

Implicit neural representations (INRs),[3–10] which model images as continuous functions rather than discrete pixel arrays, have recently gained traction for their ability to capture fine details and complex structures in data. By embedding these representations within a maximum likelihood framework, our method optimizes the INR to produce high-fidelity, speckle-free reconstructions of noisy sequences. The core of our approach is a loss function designed to predict a denoised sequence that maximizes the likelihood of observing the noisy measurements, thereby denoising the temporal scene even in the presence of motion.

Our approach significantly outperforms alternative denoising techniques for dynamic scenes with simulated speckle. Specifically, our method achieves a peak signal-to-noise ratio (PSNR) improvement of up to 4 dB when compared to alternative methods even under extreme speckle with dynamic objects. These results underscore the potential of implicit neural representations for enhancing image quality in challenging dynamic imaging scenarios.

2. PROBLEM FORMULATION

Coherent imaging systems measure the complex back-scattered field, $\tilde{y} \in \mathbb{C}^N$, from an illuminated scene. For T temporal measurements, we model this as

$$\tilde{y}_{t_i} = A_{t_i}(r_{t_i}^{1/2} \circ g_{t_i}) + n_{t_i} \quad \text{for} \quad i = 1, \ldots, T, \tag{1}$$

where $r_{t_i} \in \mathbb{R}^N$ is the time-dependent scene reflectance, $g_{t_i} \sim CN(0, I)$ is a circular Gaussian random variable representing the speckle noise, and $n_{t_i} \sim CN\left(0, \sigma^2 I\right)$ is a circular Gaussian random variable representing additive white Gaussian noise.[11,12] The measurement operator, $A_{t_i} \in \mathbb{C}^{N\times N}$, represents the physical measurement system being used. Here, we define it as

$$A_{t_i} = D^H \mathcal{D}(a)\mathcal{D}(e^{j\phi_{t_i}})D, \tag{2}$$

where $\mathcal{D}(\cdot)$ is the diagonalization operator, $a \in \mathbb{R}^N$ is the entrance-pupil transmission function, and $\phi_{t_i} \in \mathbb{R}^N$ is the phase error. $D \in \mathbb{C}^{N\times N}$ is the two dimensional discrete Fourier transform matrix scaled such that $D^H D = I$.[1,12] In this work, we explore denoising under a zero-turbulence assumption, setting $\phi_{t_i} = 0$.

3. METHODOLOGY

3.1 Maximum Likelihood-Based Loss

In order to denoise the temporal speckle noise sequence, we define a loss function for an implicit neural representation (INR) such that it approximates the scene reflectance to maximize the likelihood of the associated noisy measurements. We model the measured field $\tilde{y}$ as a multivariate Gaussian distribution

$$p(\tilde{y}_{t_i} \mid r_{t_i}) \sim CN\left(0, A_{t_i}\mathcal{D}(r_{t_i})A_{t_i}^H + \sigma^2 I\right), \tag{3}$$

with a corresponding probability density function (PDF)

$$p(\tilde{y}_{t_i} \mid r_{t_i}) = \frac{1}{\pi^k \det(\Sigma_{t_i})} \exp\left(-\tilde{y}_{t_i}^H \Sigma_{t_i}^{-1} \tilde{y}_{t_i}\right), \tag{4}$$

where k is the dimension of the complex Gaussian vector $\tilde{y}_{t_i}$, and $\det(\Sigma_{t_i})$ is the determinant of the covariance matrix $\Sigma_{t_i} = A_{t_i}\mathcal{D}(r_{t_i})A_{t_i}^H + \sigma^2 I$.[12] Given this, the negative log-likelihood over all T temporal measurements is

$$-\ln p\left(\tilde{y}_{t_1}, \tilde{y}_{t_2}, \ldots \tilde{y}_{t_T} \mid r_{t_i}\right) = \sum_{i=1}^{T} \ln\left(\det(\Sigma_{t_i})\right) + \tilde{y}_{t_i}^H \Sigma_{t_i}^{-1} \tilde{y}_{t_i}. \tag{5}$$

We parameterize Eq. (5) such that r_{t_i} is represented by an INR over space and time. We optimize that INR to minimize the negative log-likelihood, effectively learning to output the sequence of denoised measurements.

3.2 Implicit Neural Representation

We represent the time-dependent scene reflectance r_{t_i} using a multi-layer perceptron[13] (MLP)-based implicit neural representation (INR). The MLP is an eight-layer, densely connected neural network with a hidden dimension of 64, rectified linear unit (ReLU) activations, and a final sigmoid activation. The final network has 29.4K trainable parameters and is implemented using PyTorch.[14]

This INR predicts a scene reflectance pixel given unflattened spatial and temporal coordinate inputs (u, v, t_i). By evaluating the INR over all spatial and temporal coordinates of the sequence, it produces a reconstruction of the full time-dependent scene reflectance r_{t_i}. We optimize this INR using Eq. (5) to denoise the speckled temporal measurements. An illustration of this framework can be seen in Fig. 1.

4. RESULTS

We demonstrate denoising with two 32-frame sequences, one simple and the other more challenging. The first sequence is simulated motion of the Cameraman image by continuously translating a cropped portion of the image over time. This sequence simulates rigid motion (sensor moving and static scene), and it has less detail than the second sequence as well as high-contrast scene elements. The second sequence is a 30 fps video from the Need For Speed[15] dataset featuring a car drifting on a track. This sequence captures deformable motion (one object moving in otherwise static scene), with significantly more scene detail and less contrast. We simulate

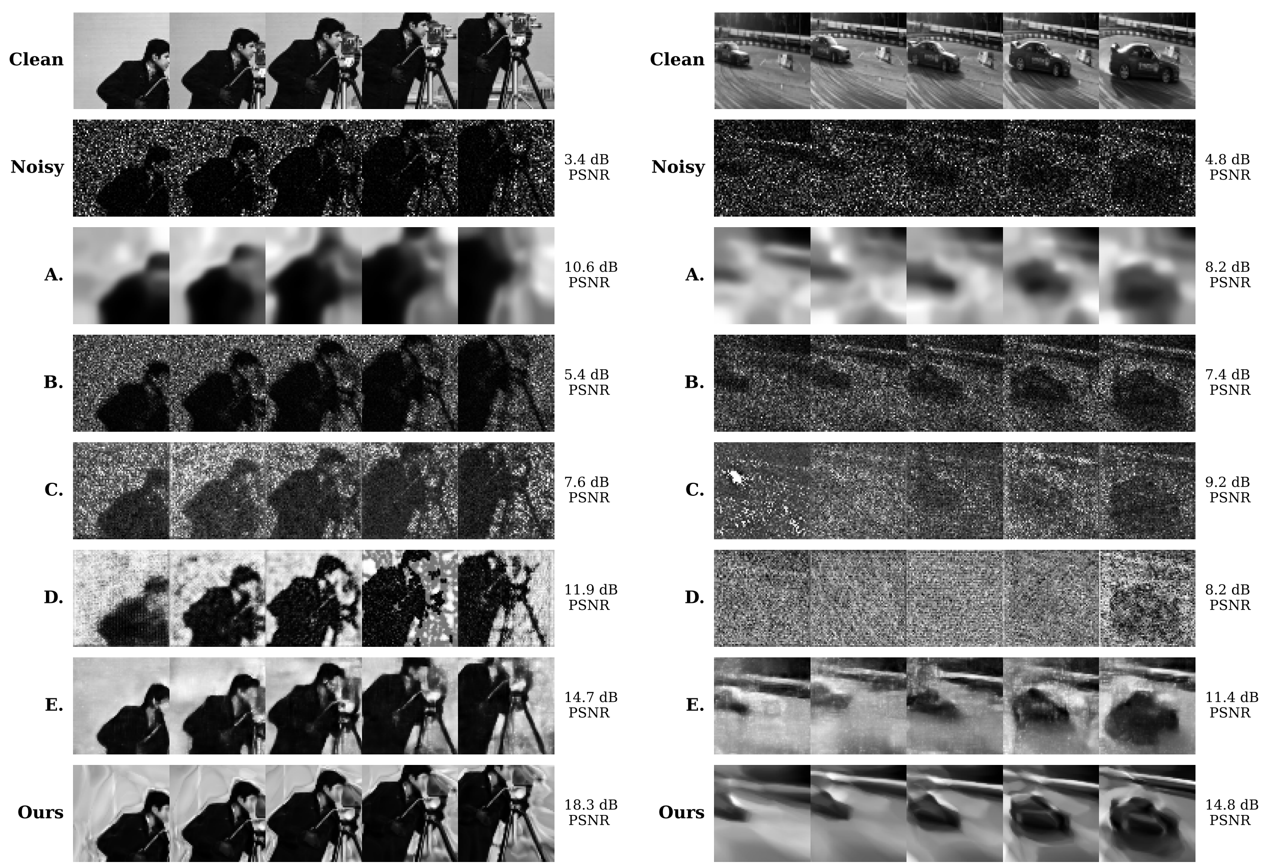

Figure 2. **Fully Resolved Speckle Denoising Performance.** Comparison of our denoising method (bottom) applied to a 64×64 resolution moving speckled sequence collected with an aperture diameter 80% the size of the sensor plane. Also shown are frame-by-frame denoising methods: A) BM3D, B) 4-Frame Averaging, C) DIP, D) DIP with our NLL loss, and E) supervised CNN speckle denoiser. Every seventh frame of the sequence is shown.

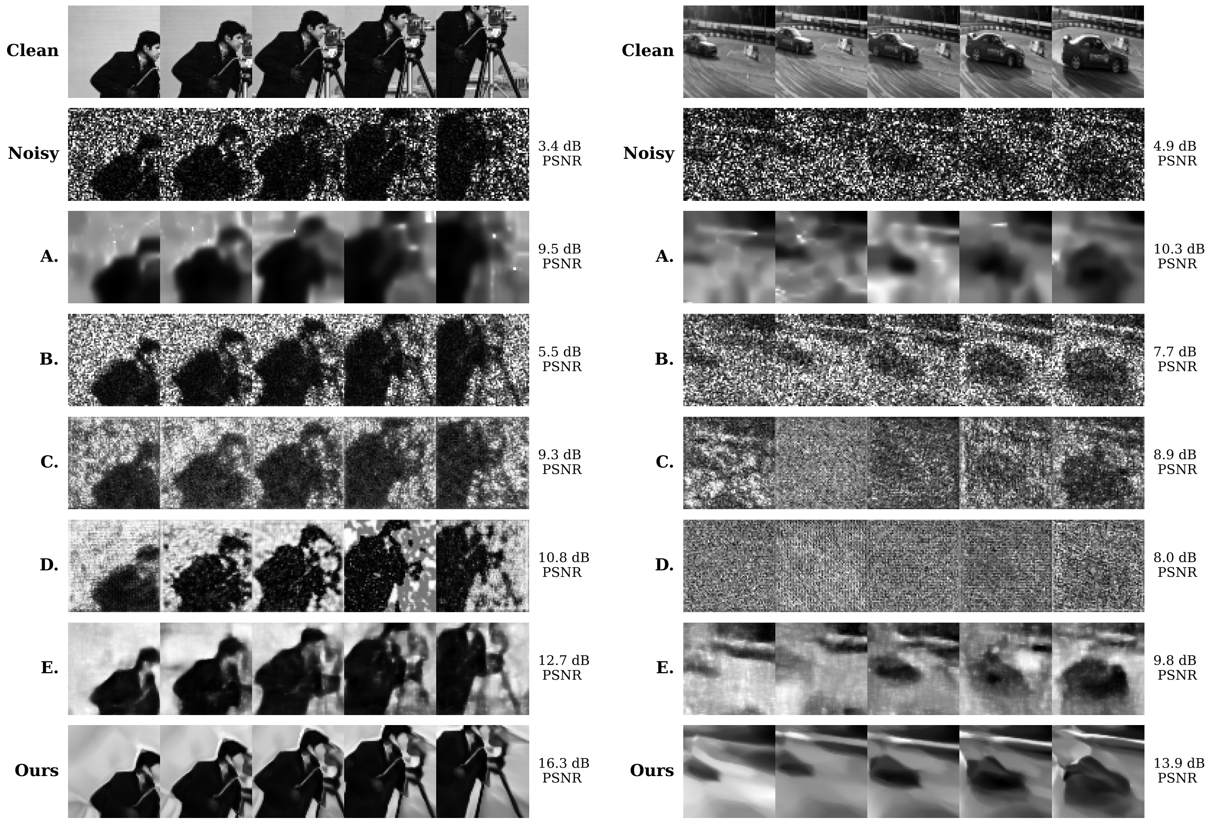

Figure 3. **Small Aperture Speckle Denoising Performance.** Comparison of our denoising method (bottom) applied to a 64×64 resolution moving speckled sequence collected with an aperture diameter 40% the size of the sensor plane (which results in blurred speckle). Also shown are frame-by-frame denoising methods: A) BM3D, B) Frame Averaging, C) DIP, D) DIP with our NLL loss, and E) supervised CNN speckle denoiser. Every seventh frame of the sequence is shown.

speckle in these sequences with variable aperture sizes. For all aperture sizes, we normalize the gain such that an equal amount of light is collected by our sensor. All sequences are resized to resolutions of 64×64 pixels.

We compare our method to five alternative denoising frameworks, all of which are applied frame-by-frame to the sequence. These include Block-Matching and 3D Filtering (BM3D),[16] frame averaging (over four neighboring frames), Deep Image Prior (DIP),[17] DIP utilizing our negative log likelilood loss function, and a supervised CNN speckle denoiser similar to Pellizzari et al.[18,19] The CNN-based denoiser is a 31M parameter model trained on a dataset of augmented images from the DIV2K[20] dataset using our forward model to simulate speckle noise. We train a new model for each aperture size tested.

With larger aperture sizes, we demonstrate an average 3.6 and 3.4 dB PSNR improvement on the Cameraman and car sequences, respectively, when compared to the supervised CNN denoiser (the best baseline). Our method retains significantly more detail and temporal consistency than the baselines, as shown in Fig. 2. For smaller apertures, the speckle noise grows more spatially correlated which makes the problem more challenging. Our method achieves an average 3.6 and 4.1 dB PSNR improvement on both sequences. Though less detail is captured—especially in the car sequence—our method is still able to emphasize major elements of the scene and accurately reproduce motion, shown in Fig. 3. Notably, our method significantly outperforms DIP modified to use our loss function, indicating that the INR framework is essential for improved performance.

Our method's primary weakness is its computational cost, which stems from the costly matrix inversion embedded within our loss function. As an unsupervised method, our approach must be fit to each video sequence independently. Accordingly, to reconstruct a 64×64 resolution 32 frame sequence, our unoptimized implementation requires requires 82 minutes for optimal convergence on an NVIDIA RTX 4080. For comparison, the CNN denoiser must be trained for each aperture size for 216 minutes, but can then rapidly perform inference in near real-time.

5. CONCLUSION

In this paper, we introduce a novel approach for speckle denoising in dynamic scenes using time-varying implicit neural representations (INRs). Our unsupervised method leverages the continuous modeling capabilities of INRs within a maximum likelihood framework to achieve high-fidelity reconstructions of noisy temporal sequences. We demonstrate our method on temporal sequences featuring rigid and deformable motion with simulated speckle. Our method outperformed alternative denoising techniques, achieving up to a 4 dB improvement in peak signal-to-noise ratio (PSNR). However, our method is computationally expensive—its unsupervised nature does not scale as well as a supervised alternative for long temporal sequences.

ACKNOWLEDGMENTS

The authors thank Profs. Shirin Jalali, Arian Maleki, and Casey Pellizzari for their input on implementation methodology and evaluation framework. This work was supported in part by ARO ECP award no. W911NF-24-2-0113, AFOSR Young Investigator Program award no. FA9550-22-1-0208, NSF CAREER award no. 2339616, ONR award no. N00014-23-1-2752, and a seed grant from SAAB, Inc.

The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Dept. of Army or the Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) or the U.S. Government.

REFERENCES

[1] Goodman, J. W., [*Speckle Phenomena in Optics: Theory and Applications*], Roberts and Company Publishers (2007).

[2] Kumar, M., Tounsi, Y., Kaur, K., Nassim, A., Mandoza-Santoyo, F., and Matoba, O., "Speckle denoising techniques in imaging systems," *J. Opt.* **22**, 063001 (June 2020).

[3] Park, J. J., Florence, P., Straub, J., Newcombe, R., and Lovegrove, S., "DeepSDF: Learning continuous signed distance functions for shape representation," in [*2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], 165–174, IEEE (June 2019).

[4] Sitzmann, V., Martel, J., Bergman, A., Lindell, D., and Wetzstein, G., "Implicit neural representations with periodic activation functions," *Adv. Neural Inf. Process. Syst.* **33**, 7462–7473 (2020).

[5] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R., "NeRF: representing scenes as neural radiance fields for view synthesis," *Commun. ACM* **65**, 99–106 (Jan. 2022).

[6] Liu, J., Balaji, M. M., Metzler, C. A., Asif, M. S., and Rangarajan, P., "Solving inverse problems using self-supervised deep neural nets," in [*Computational Optical Sensing and Imaging*], CTh5A–2, Optica Publishing Group (2021).

[7] Liu, R., Sun, Y., Zhu, J., Tian, L., and Kamilov, U. S., "Recovery of continuous 3d refractive index maps from discrete intensity-only measurements using neural fields," *Nature Machine Intelligence* **4**(9), 781–791 (2022).

[8] Feng, B. Y., Guo, H., Xie, M., Boominathan, V., Sharma, M. K., Veeraraghavan, A., and Metzler, C. A., "NeuWS: Neural wavefront shaping for guidestar-free imaging through static and dynamic scattering media," *Sci. Adv.* **9** (June 2023).

[9] Zhou, H., Feng, B. Y., Guo, H., Lin, S., Liang, M., Metzler, C. A., and Yang, C., "Fourier ptychographic microscopy image stack reconstruction using implicit neural representations," *Optica* **10**(12), 1679–1687 (2023).

[10] Qadri, M., Zhang, K., Hinduja, A., Kaess, M., Pediredla, A., and Metzler, C. A., "Aoneus: A neural rendering framework for acoustic-optical sensor fusion," in [*ACM SIGGRAPH 2024 Conference Papers*], 1–12 (2024).

[11] Goodman, J. W., [*Introduction to Fourier Optics*], W.H. Freeman, New York, NY, 3 ed. (Jan. 2005).

[12] Pellizzari, C. J., Spencer, M. F., and Bouman, C. A., "Phase-error estimation and image reconstruction from digital-holography data using a bayesian framework," *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **34**, 1659–1669 (Sept. 2017).

[13] Murtagh, F., "Multilayer perceptrons for classification and regression," *Neurocomputing* **2**, 183–197 (July 1991).

[14] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S., "PyTorch: An imperative style, high-performance deep learning library," *Adv. Neural Inf. Process. Syst.* **32** (2019).

[15] Galoogahi, H. K., Fagg, A., Huang, C., Ramanan, D., and Lucey, S., "Need for speed: A benchmark for higher frame rate object tracking," *ICCV* , 1134–1143 (Mar. 2017).

[16] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K., "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.* **16**, 2080–2095 (Aug. 2007).

[17] Lempitsky, V., Vedaldi, A., and Ulyanov, D., "Deep image prior," in [*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 9446–9454, IEEE (June 2018).

[18] Pellizzari, C. J., Bate, T. J., Donnelly, K. P., Buzzard, G. T., Bouman, C. A., and Spencer, M. F., "Coherent plug-and-play artifact removal: Physics-based deep learning for imaging through aberrations," *Opt. Lasers Eng.* **164**, 107496 (May 2023).

[19] Pellizzari, C. J., Bate, T. J., Mandyam, M. G., Radosevich, C. J., Horst, S., and Spencer, M. F., "Speckle-free coherent imaging through deep turbulence," *Optics Letters* **49**, 3508–3511 (June 2024).

[20] Agustsson, E. and Timofte, R., "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in [*The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*], (July 2017).