

PAPER • OPEN ACCESS

Prediction of the morphological evolution of a splashing drop using an encoder–decoder

To cite this article: Jingzu Yee *et al* 2023 *Mach. Learn.: Sci. Technol.* **4** 025002

View the [article online](#) for updates and enhancements.

You may also like

- [Drop impact on substrates with heterogeneous stiffness](#)
Yang Cheng, , Jian-Gen Zheng et al.
- [Visualization and mechanisms of splashing erosion of electrodes in a DC air arc](#)
Yi Wu, Yufei Cui, Mingzhe Rong et al.
- [Stability and erosion of melt layers developed on plasma facing components of tokamaks](#)
G. Miloshevsky and A. Hassanein



PAPER

OPEN ACCESS

RECEIVED
14 December 2022REVISED
6 March 2023ACCEPTED FOR PUBLICATION
23 March 2023PUBLISHED
6 April 2023

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Prediction of the morphological evolution of a splashing drop using an encoder–decoder

Jingzu Yee¹ , Daichi Igarashi(五十嵐大地)¹, Shun Miyatake(宮武駿)¹
and Yoshiyuki Tagawa(田川義之)^{1,2,*} 

¹ Department of Mechanical Systems Engineering, Tokyo University of Agriculture and Technology, Koganei, Tokyo 184-8588, Japan

² Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Koganei, Tokyo 184-8588, Japan

* Author to whom any correspondence should be addressed.

E-mail: tagawayo@cc.tuat.ac.jp

Keywords: multiphase flow, drop impact, splashing, machine learning, computer vision

Abstract

The impact of a drop on a solid surface is an important phenomenon that has various implications and applications. However, the multiphase nature of this phenomenon causes complications in the prediction of its morphological evolution, especially when the drop splashes. While most machine-learning-based drop-impact studies have centred around physical parameters, this study used a computer-vision strategy by training an encoder–decoder to predict the drop morphologies using image data. Herein, we show that this trained encoder–decoder is able to successfully generate videos that show the morphologies of splashing and non-splashing drops. Remarkably, in each frame of these generated videos, the spreading diameter of the drop was found to be in good agreement with that of the actual videos. Moreover, there was also a high accuracy in splashing/non-splashing prediction. These findings demonstrate the ability of the trained encoder–decoder to generate videos that can accurately represent the drop morphologies. This approach provides a faster and cheaper alternative to experimental and numerical studies.

1. Introduction

‘Constant dripping wears through the stone’ is a famous Chinese idiom that means ‘perseverance yields success’. This idiom perfectly describes the phenomenon of the impact of a drop on a solid surface, which might appear to be insignificant but actually has great implications and many applications in both nature and industry [1–5]. This is especially true when splashing occurs, i.e. when the impacting drop breaks up and ejects secondary droplets [6–10] instead of just spreading over the surface until it reaches a maximum diameter [11–15]. To name just a few of the implications of splashing, in nature, it is the main cause of erosion and the propagation of contaminants, while in industry, it can cause visible decreases in printing and paint quality [16–18].

To minimise the adverse effects of this fascinating and beautiful phenomenon, many studies have been performed to understand the mechanisms and model the morphology of a splashing drop using first-principles approaches, including theoretical and numerical analyses [19–22]. Because of the multiphase nature of a drop impact, the occurrence of splashing is heavily influenced by the physical parameters involved. These can be categorized into the impact conditions and the respective physical properties of the liquid drop, the solid surface, and the ambient air [23–28]. Most often, the relationships between the occurrence of splashing and each of these parameters are not simply monotonic, but rather there is an interplay between two or more of them. Sometimes, a given parameter can either promote or suppress splashing depending on other parameters [29–31]. Furthermore, splashing can happen in different ways, including as a prompt splash, corona splash, receding breakup, and magic carpet breakup [8–10]. For these reasons, the morphological study of the splashing phenomenon can present complications.

In recent years, with the increasing availability and accessibility of data, data-driven approaches—and machine learning in particular—have attracted increasing attention among fluid researchers as a faster and cheaper alternative or complement to experimental and numerical studies [32–39]. Regarding drop impacts, several machine-learning-based studies have been carried out [40–43]. Notably, a number of studies on predicting the maximum spreading factor of a non-splashing drop under various conditions were published in 2022. For example, Yancheshme *et al* used the random forest to predict the maximum spreading of a drop on hydrophobic and superhydrophobic surfaces [40]. Also, Tembely *et al* compared the performances of the linear regression model, decision tree, random forest, and gradient boost regression model on predicting the maximum spreading of a drop on surfaces of various wettabilities [41]. Even for droplet-particle collisions, Yoon *et al* used a multi-layer feedforward neural network (FNN) to predict the maximum spreading diameter under a significantly wide range of impact conditions [42]. Other than the maximum spreading diameter, such physical parameters-based machine learning was also applied for the investigations of the drop impact force and the splashing mechanisms. Dickerson *et al* used a physical parameters-based ensemble learning algorithm to predict the drop impact force on concave targets [44]. As for splashing mechanisms, Pierzyna *et al* [43] successfully improved the splashing threshold proposed by Riboux and Gordillo [7]. Although the machine learning models in these studies showed excellent performances, they were designed to use physical parameters as inputs and outputs.

As the next step, the objective of this study was to predict the evolving dynamics of the morphology of a splashing drop during the impact. For that, we focused on the application of image data to explore the possibility of using computer vision—i.e. the ability of machine learning to process and generate images—as a strategy for tackling the complications of predicting the morphology of a splashing drop. This was inspired by Yee *et al*, who attempted to understand the splashing mechanism through the extraction of morphological features in splashing drops by visualising how a computer vision-based FNN classified splashing and non-splashing drops [45]. In this study, the machine-learning algorithm of an encoder–decoder [46–52] was trained to generate videos in the form of image sequences that can accurately represent the actual morphological evolutions of splashing and non-splashing drops during their impact on a solid surface under specific physical parameters. During the training of the encoder–decoder, interpolation prediction was performed by withholding intermediate data points. The purpose of this is to build an encoder–decoder that can generate accurate post-impact image sequences for the intermediate data points. This will provide a faster and cheaper alternative to experiments or numerical simulations on new data points.

The remainder of this paper is structured as follows. The dataset used in this study are described in detail in section 2 and the detailed implementations of the encoder–decoder model in section 3. The results and discussion are presented in section 4: the post-impact image sequences generated by the trained encoder–decoder are qualitatively evaluated in section 4.1, and they are quantitatively evaluated based on the spreading diameter in section 4.2 and the accuracy of splashing/non-splashing prediction in section 4.3. Analysis on the prediction process of the trained encoder-decoder is explained in section 4.4. A conclusion summarizing the results is presented in section 5.

2. Dataset

2.1. Data collection

To collect the image data, experiments were performed using the setup shown in figure 1. A syringe was used to supply liquid ethanol (Hayashi Pure Chemical Ind., Ltd; density $\rho = 789 \text{ kg m}^{-3}$, surface tension $\gamma = 2.2 \times 10^{-2} \text{ N/m}$, and dynamic viscosity $\mu = 10^{-3} \text{ Pa}\cdot\text{s}$) to form a drop of area-equivalent diameter $D_0 = (2.59 \pm 0.08) \times 10^{-3} \text{ m}$ at a needle (internal diameter $0.97 \times 10^{-3} \text{ m}$), where it detached and impacted the hydrophilic surface of a glass substrate (Muto Pure Chemicals Co., Ltd star frost slide glass 511 611); the free-fall distance of the drop H (0.04–0.60 m) was the only manipulated variable. The measured impact velocity U_0 was in the range $0.82\text{--}3.18 \text{ m s}^{-1}$, and the resulting Weber number We , which was computed using

$$We = \rho U_0^2 D_0 / \gamma, \quad (1)$$

ranged from 63 to 947. For each drop impact, a video was captured using a high-speed camera (Photron, FASTCAM SA-X) at a frame rate of $45\,000 \text{ s}^{-1}$ and a spatial resolution of $(1.46 \pm 0.02) \times 10^{-5} \text{ m pixel}^{-1}$. After a manual frame-by-frame inspection for the presence of secondary droplets, each of the videos was labelled according to the outcome: splashing or non-splashing. A total of 249 videos were recorded: 141 splashing and 108 non-splashing.

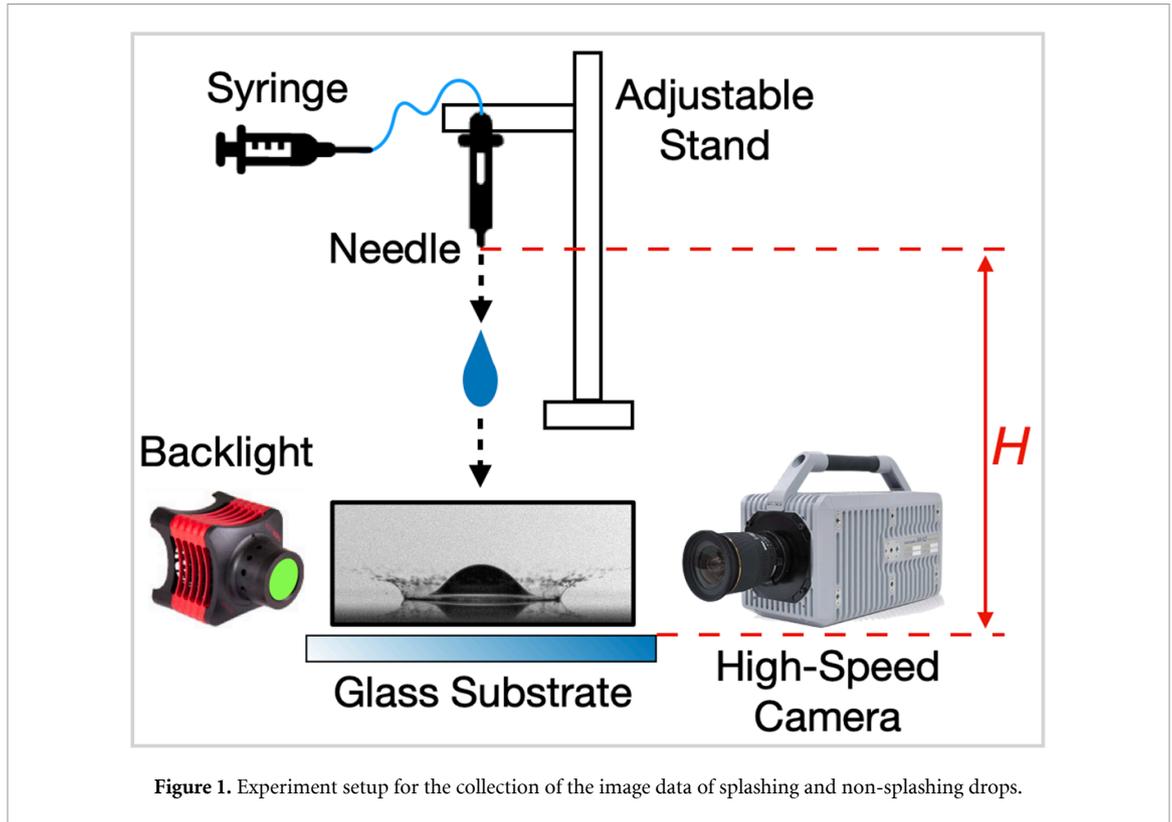


Figure 1. Experiment setup for the collection of the image data of splashing and non-splashing drops.

To evaluate the collected data in terms of the We and the D_0 , figure 2 was plotted. Figure 2(a) shows the plot of We versus H of the collected data. The blue and green circles show the splashing and non-splashing data, respectively. As shown by the areas filled with grey, the splashing thresholds in terms of impact height and Weber number were $H = 0.20$ m and $We = 348$, respectively. Note that some impacting drops with H or We values greater than or equal to the splashing thresholds did not splash. The highest values of H and We for a non-splashing drop were $H = 0.22$ m and $We = 386$. Hence, there is a splashing transition within $0.20 \text{ m} \leq H \leq 0.22 \text{ m}$ or $348 \leq We \leq 386$. The black dashed-dotted line shows the plot of the theoretical Weber number We_{theo} against H . The equation used to compute We_{theo} is

$$We_{\text{theo}} = \rho U_{\text{theo}}^2 D_0 / \gamma = 2\rho D_0 g H / \gamma, \quad (2)$$

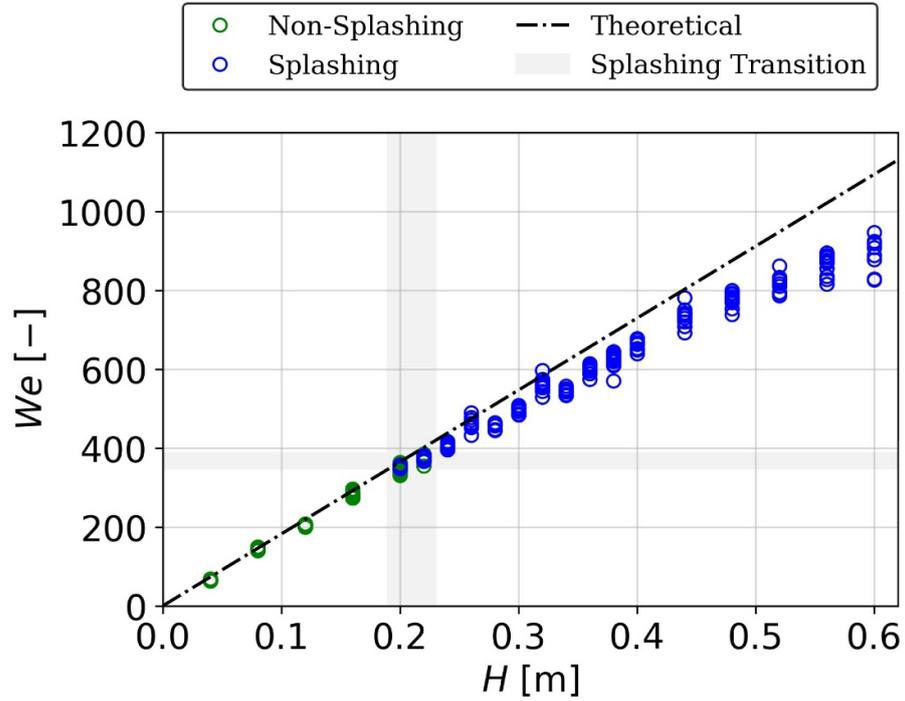
where $U_{\text{theo}} (= \sqrt{2gH})$ is the theoretical free-falling velocity and the gravitational acceleration $g = 9.81 \text{ m s}^{-2}$. The collected data shows the same linear trend as the theory shown by the black dashed-dotted line. Note that the deviation from the theory increases with H because the drag increased with H causing U_0 to deviate from U_{theo} .

As for evaluation on the D_0 , the plot versus H of the collected data is shown in figure 2(b). The black and red dashed lines show the mean value and the range of two times the standard deviation, respectively. As shown in the figure, D_0 of all splashing and non-splashing data lie between the range of two times the standard deviation, indicating that there was no significant bias in terms of the drop size.

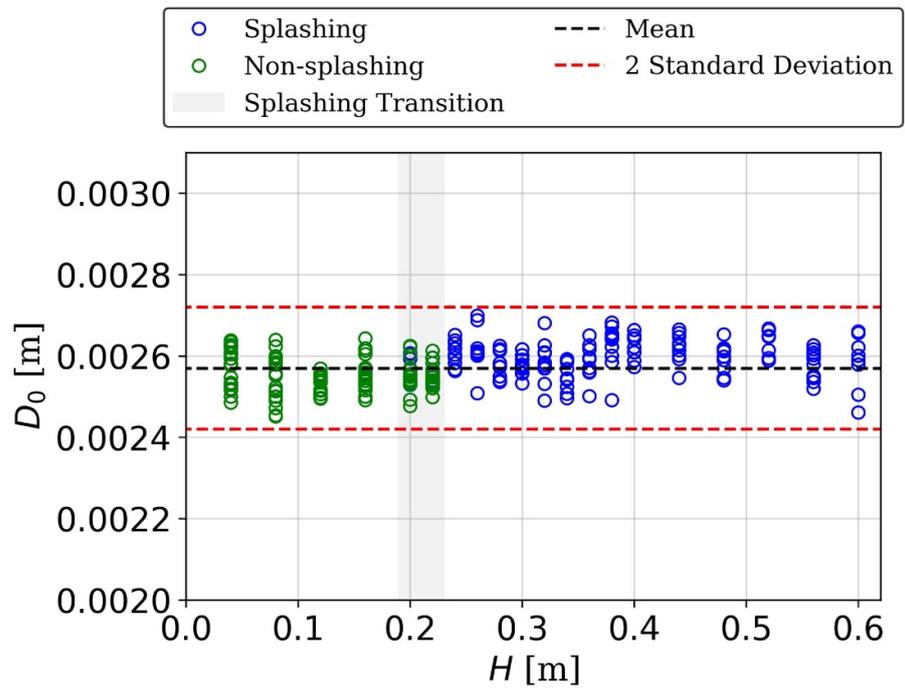
2.2. Digital image processing

Digital image processing was performed using an in-house MATLAB code to detect the drop and the glass substrate in each frame of the recorded high-speed videos. The main image processing toolbox used for the detection is region boundary tracing [53]. The detection was enhanced using other toolboxes, such as binarisation [54], non-local means filtering [55], circle detection [56, 57], and edge detection [58]. Detection of the drop and the substrate was necessary for measuring the area-equivalent diameter D_0 and the impact velocity U_0 of each data. More importantly, it also prepared the pre-impact and post-impact image sequences for the input and the output, respectively, of the encoder-decoder model.

The computation of D_0 and U_0 was performed as follows. First, the frame when the drop first touched the substrate ($t = 0$) is identified. In the code, it was defined as the frame when the lowest point of the drop is



(a)



(b)

Figure 2. (a) Weber number We and (b) area-equivalent diameter D_0 versus the impact height H of the collected data.

lower than or equal to the highest point of the glass substrate. The illustration is shown in figure 3. After the frame $t = 0$ was identified, the area bounded by the drop was detected by using the toolbox for region boundary tracing to compute D_0 and the centre of mass. The same procedure was applied to the previous nine frames to compute the average D_0 and U_0 , which were used to compute We . The related results are shown and discussed in section 2.1.

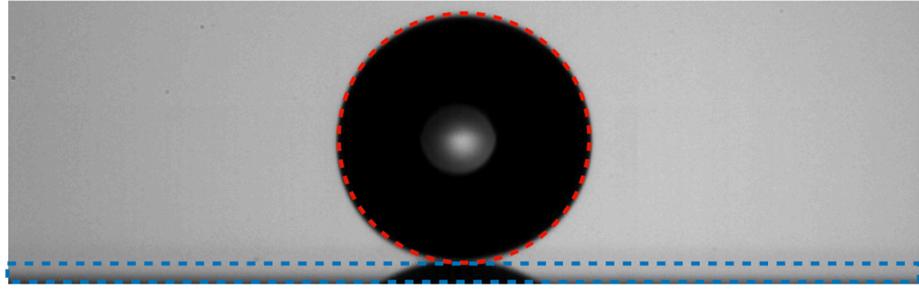


Figure 3. Illustration of the image processing performed on the collected data to identify the frame when the drop first touched the substrate ($t = 0$). The red dashed and the blue dashed lines show the area bound by the drop and the glass substrate, respectively, as detected by the in-house MATLAB code through region boundary tracing. This frame was identified as $t = 0$ because the lowest point of the red dashed line is equal to the highest point of the blue dashed line.

t	$H = 0.08$ m; $We = 145$	$H = 0.16$ m; $We = 277$	$H = 0.24$ m; $We = 398$	$H = 0.56$ m; $We = 871$
- 2 frames				
- 1 frame				
0				

Figure 4. Examples of the input image sequences.

An input image sequence should indirectly provide the physical parameters of the drop impact to the encoder–decoder for the prediction of the post-impact morphologies of a drop. Therefore, an input image sequence was prepared as follows. First, similar to the computation of D_0 and U_0 , the frame when the drop first touched the substrate ($t = 0$) was identified. The frame was then cropped from $288 \times 1024 \text{ pixel}^2$ to $250 \times 250 \text{ pixel}^2$, with the drop at the centre of the cropped image while the surface of the glass substrate at the bottom of the cropped image. This was to ensure the high similarity of the frames extracted for all the collected data. The same procedure was applied to the two preceding frames. Finally, the cropped frames of $t = 0$ and the two preceding frames were combined to form an input image sequence. Some examples of the input image sequences are shown in figure 4. Note that all the videos were recorded at the same frame rate, thus the positions of the drop in the first two frames of a pre-impact image sequence change according to We , where a drop with a higher We had higher positions in the first two frames due to the higher U_0 . This indirectly provided the encoder–decoder the information about the We of a drop.

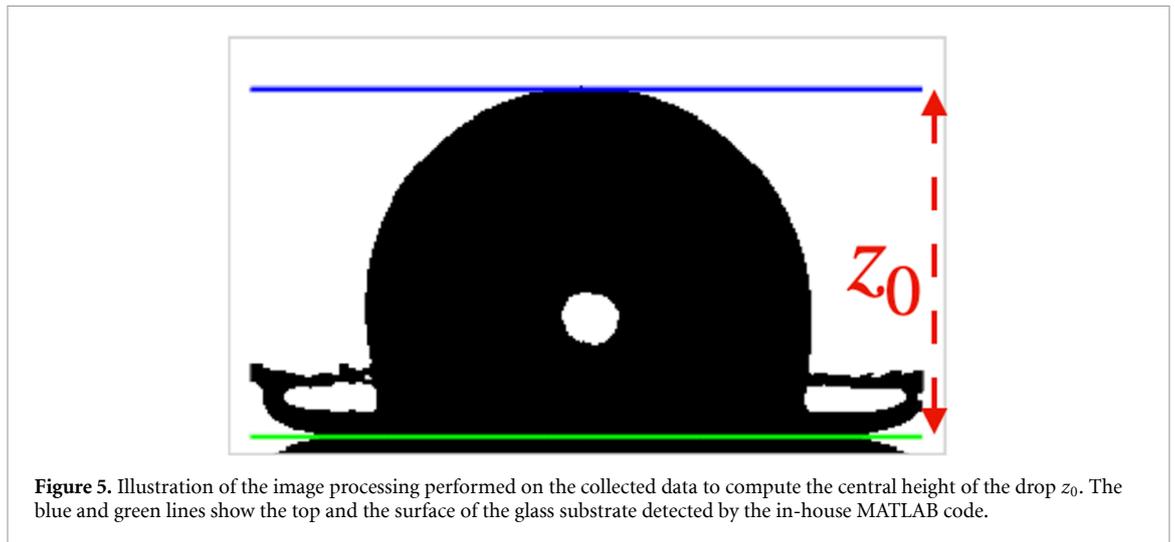


Figure 5. Illustration of the image processing performed on the collected data to compute the central height of the drop z_0 . The blue and green lines show the top and the surface of the glass substrate detected by the in-house MATLAB code.

z_0/D_0	$H = 0.08$ m; $We = 145$	$H = 0.16$ m; $We = 277$	$H = 0.24$ m; $We = 398$	$H = 0.56$ m; $We = 871$
0.875				
0.750				
0.625				
0.500				
0.375				
0.250				
0.125				

Figure 6. Examples of the output image sequences.

As for the output image sequence, it should show the morphologies of the drop throughout the impact. First, the top and the bottom of the drop during the impact were identified to compute the central height z_0 . The illustration is shown in figure 5. The frames when the normalised central height of the drop $z_0/D_0 = 0.875, 0.750, 0.625, 0.500, 0.375, 0.250,$ and 0.125 were identified and extracted. Since D_0 and the spatial resolutions were kept, z_0 in each of these frames had the same value for all collected data. According to Lagubeau *et al* [21], these values of z_0/D_0 correspond to the pressure impact and self-similar inertial regimes. Each of these seven frames was then cropped from $288 \times 1024 \text{ pixel}^2$ to $640 \times 200 \text{ pixel}^2$, with the drop at the centre of the cropped image while the surface of the glass substrate at the bottom of the cropped image. Similar to the input image sequences, this was to ensure the high similarity of the frames extracted for all the collected data. Finally, these seven cropped frames were binarised and combined to form an output image sequence. Some examples of the output image sequences are shown in figure 6. Note that binarisation was performed so that each pixel occupied by the drop morphology had an intensity value of 0 while each pixel not occupied had an intensity value of 1. Thus, the image-generation process can be considered as a pixel-by-pixel prediction of whether a pixel is occupied by the drop morphology. Such a strategy was inspired by volume of fluid method, a commonly used free-surface modelling technique [59].

2.3. Data segmentation

A well-trained encoder–decoder that can generate accurate post-impact image sequences for the intermediate data points can provide a faster and cheaper alternative to experiments or numerical simulations on new data points. To examine the ability of the trained encoder–decoder to perform interpolation prediction, all image sequences of drops with $H = 0.08, 0.16, 0.24, 0.32, 0.40, 0.48,$ and 0.56 m were excluded from the training and reserved only for the testing. For the remaining image sequences with $H = 0.04, 0.12, 0.20, 0.22, 0.26, 0.28, 0.30, 0.34, 0.36, 0.38, 0.44, 0.52,$ and 0.60 m, 90%–10% segmentation was performed according to each H value. Thus, 146 sets (about 59%) and 103 sets (about 41%) were used for training and testing, respectively. Hereafter, the reserved H and the non-reserved H values are referred to as the interpolation impact height H and training impact height H , respectively.

3. Encoder–decoder

An encoder–decoder is a type of artificial neural network that can encode higher-dimensional data into lower-dimensional representation and subsequently decode the encoded representation into the desired output of higher dimension [46]. It is widely used for dimensionality reduction and data generation both in fundamental research and practical applications. An example of its applications for fundamental research is the study by Lu *et al* that designed DeepONet based on an encoder–decoder to learn diverse linear/nonlinear explicit and implicit operators [47]. For the discovery of novel materials, Krenn *et al* used a variational autoencoder to reconstruct the molecules [48]. Also, Kämring *et al* applied unsupervised machine learning using an autoencoder to predict the phase transitions from the experimental data of a Haldane-like model realised with ultracold atoms [49]. On the other hand, for practical applications, Badrinarayanan designed SegNet based on a convolutional encoder–decoder for image segmentation [50]. In addition, Khan *et al* effectively utilised the framework of an encoder–decoder to extract building footprints from aerial images [51] and to predict COVID-19 hotspots [52]. Especially, for the prediction of COVID-19 hotspots, they successfully incorporated a hybrid model of convolutional neural network and long short-term memory to process time-series data.

3.1. Architecture

In this study, an encoder–decoder with the optimised architecture shown in figure 7 was trained to generate a post-impact image sequence as the output, from a pre-impact image sequence as the input. This was implemented in the Python programming language on the Jupyter Notebook [60] using the TensorFlow libraries [61].

As mentioned in section 2.2, each frame of an input image sequence cropped into size $h_{\text{in}} \times w_{\text{in}} = 250 \times 250$ pixels². With three frames, an input image sequence had the shape of $h_{\text{in}} \times w_{\text{in}} \times f_{\text{in}} = 250 \times 250 \times 3$, where f_{in} is the number of frames in an input image sequence. As for an output image sequence, each frame was cropped into size $h_{\text{in}} \times w_{\text{in}} = 640 \times 200$ pixels². With seven frames, an output image sequence had the shape of $h_{\text{out}} \times w_{\text{out}} \times f_{\text{out}} = 640 \times 200 \times 7$, where f_{out} is the number of frames in an output image sequence.

In the encoder, an input image sequence is first flattened into a one-dimensional column vector: $\mathbf{X} \in \mathbb{R}^{h_{\text{in}} \times w_{\text{in}} \times f_{\text{in}}} \rightarrow \mathbf{s}_{\text{in}} \in \mathbb{R}^{N_{\text{in}}}$, for $N_{\text{in}} = 250 \times 250 \times 3 = 187,500$. Then, it is encoded into the latent space by a linear function and a leaky rectified linear unit (ReLU) activation function. The linear function is

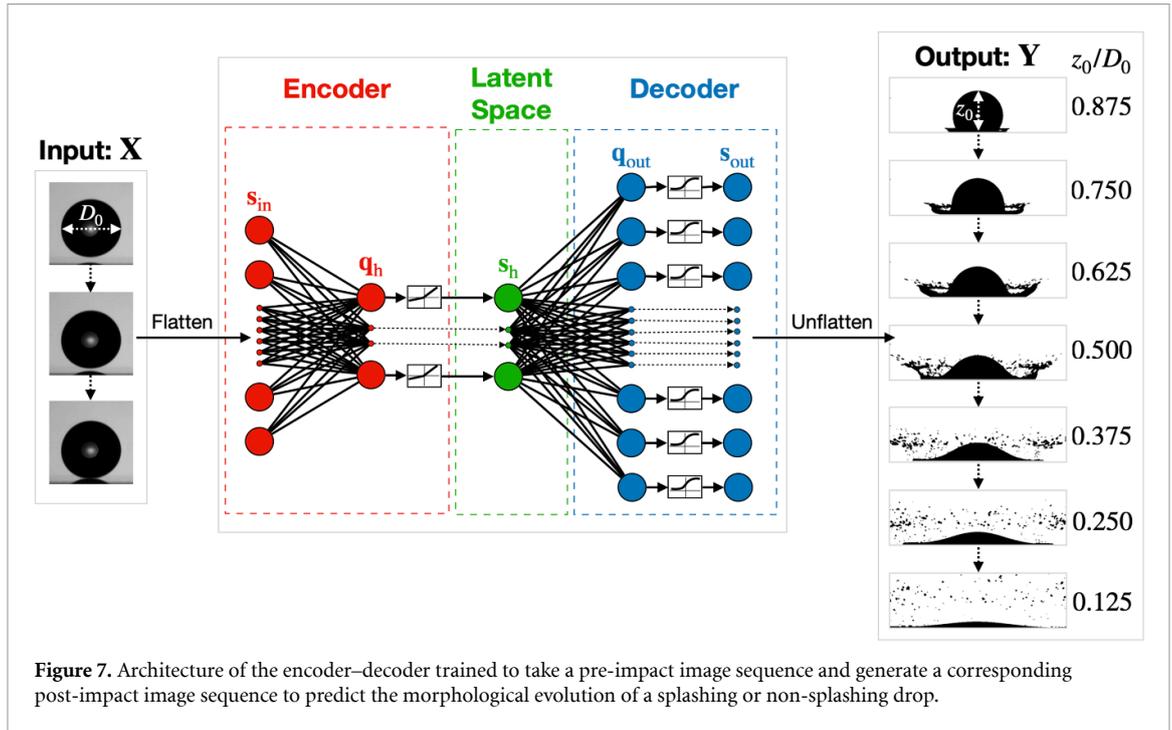
$$\mathbf{q}_{\text{h}} = \mathbf{W}_{\text{enc}} \mathbf{s}_{\text{in}} + \mathbf{b}_{\text{enc}}, \quad (3)$$

where $\mathbf{q}_{\text{h}} \in \mathbb{R}^{N_{\text{enc}}}$ is the output vector of the linear equation, $\mathbf{W}_{\text{enc}} \in \mathbb{R}^{N_{\text{enc}} \times N_{\text{in}}}$ is the encoding weight matrix, and $\mathbf{b}_{\text{enc}} \in \mathbb{R}^{N_{\text{enc}}}$ is the encoding bias vector. Here, N_{enc} is the number of encoded elements, which is set as 32 through hyperparameter tuning. The leaky ReLU function is

$$s_{\text{h},i} = \begin{cases} q_{\text{h},i} & \text{for } q_{\text{h},i} \geq 0 \\ \alpha q_{\text{h},i} & \text{for } q_{\text{h},i} < 0 \end{cases} \quad (4)$$

for $i = 1, \dots, N_{\text{enc}}$, where the output of the function $s_{\text{h},i}$ is an element of the encoded vector in the latent space $\mathbf{s}_{\text{h}} \in \mathbb{R}^{N_{\text{enc}}}$ and α is the negative-slope coefficient, which is set as 0.3.

In the decoder, $\mathbf{s}_{\text{h}} \in \mathbb{R}^{N_{\text{enc}}}$ is decoded into the output layer by a linear function and a sigmoid activation function. The linear function is



$$\mathbf{q}_{out} = \mathbf{W}_{dec}\mathbf{s}_h + \mathbf{b}_{dec}, \quad (5)$$

where $\mathbf{q}_{out} \in \mathbb{R}^{N_{dec}}$ is the output vector of the linear equation, $\mathbf{W}_{dec} \in \mathbb{R}^{N_{dec} \times N_{enc}}$ is the decoding weight matrix, and $\mathbf{b}_{dec} \in \mathbb{R}^{N_{dec}}$ is the decoding bias vector. Here, N_{dec} is the number of decoded elements, which is the total number of pixels in a post-impact image sequence. Thus, $N_{dec} = h_{out} \times w_{out} \times f_{out} = 640 \times 200 \times 7 = 896\,000$. The sigmoid function is

$$s_{out,i} = \sigma(q_{out,i}) = \frac{1}{1 + e^{-q_{out,i}}}, \quad (6)$$

for $i = 1, \dots, N_{dec}$, where the output of the function $s_{out,i}$ is an element of the one-dimensional column vector in the output layer $\mathbf{s}_{out} \in \mathbb{R}^{N_{dec}}$. This function saturates negative and positive values to 0 and 1, respectively, to generate a binary number in each pixel. Finally, the vector \mathbf{s}_{out} is unflattened to form the post-impact image sequences: $\mathbf{s}_{out} \in \mathbb{R}^{N_{dec}} \rightarrow \mathbf{Y} \in \mathbb{R}^{h_{out} \times w_{out} \times f_{out}}$.

The prediction by the encoder–decoder can be defined as a mapping $\mathbf{Y} = F(\mathbf{X}, \theta)$. Here, θ denotes the mapping parameters that consist of all the elements of weight matrices and bias vectors: $\mathbf{W}_{enc} \in \mathbb{R}^{N_{enc} \times N_{in}}$, $\mathbf{W}_{dec} \in \mathbb{R}^{N_{dec} \times N_{enc}}$, $\mathbf{b}_{enc} \in \mathbb{R}^{N_{enc}}$, and $\mathbf{b}_{dec} \in \mathbb{R}^{N_{dec}}$. Therefore, the total number of parameters N_p can be calculated using the following equation:

$$N_p = N_{enc} \times N_{in} + N_{dec} \times N_{enc} + N_{enc} + N_{dec} \quad (7)$$

$$= N_{enc}(N_{in} + N_{dec} + 1) + N_{dec}, \quad (8)$$

where $N_p = 35\,568\,032$.

3.2. Training

The training of the encoder–decoder was performed by minimising a binary cross-entropy loss function. The binary cross-entropy loss function was chosen because this study considers the image-generation process as a pixel-by-pixel prediction of whether a pixel was occupied by the drop morphology. The loss l was computed using the following equation:

$$l(\mathbf{s}_{true}, \mathbf{s}_{out}) = \sum_{i=1}^{N_{dec}} [-s_{true,i} \ln(s_{out,i}) - (1 - s_{true,i}) \ln(1 - s_{out,i})], \quad (9)$$

for $i = 1, \dots, N_{\text{dec}}$, where \mathbf{s}_{true} is the one-dimensional column vectors of the actual post-impact image sequences. When \mathbf{s}_{out} is close to \mathbf{s}_{true} , l will be close to 0. On the contrary, when \mathbf{s}_{out} is not equal to \mathbf{s}_{true} , l increases dramatically as \mathbf{s}_{out} deviates further from \mathbf{s}_{true} .

To determine how each element of the weight matrices and the bias vectors should be tweaked to minimise l , a backpropagation algorithm was applied to compute the gradient of the computed l with respect to each element. For the optimization, the adaptive moment (Adam) optimizer [62] was used to effectively tweak these elements in the direction of descending gradients.

3.3. Hyperparameter tuning

Hyperparameter tuning was performed concurrently with the training of the encoder–decoder. The goal was to find the best architecture that can achieve the desired performance with the lowest computational cost possible.

In terms of desired performance, the encoder–decoder should be able to generate the morphologies of non-splashing drops and splashing drops of different physical parameters. In the case when the desired performance is not achieved, the encoder–decoder only generates the morphologies of non-splashing drops regardless of the input.

On the other hand, a low computational cost could be achieved by a simple architecture that has a small N_p . For that, the number of hidden layers in the encoder and the decoder was reduced to zero. As for N_{enc} , it was reduced to 32 by using the leaky ReLU function instead of the ReLU function for the activation of the encoded elements. Note that the α of the ReLU function is 0. The encoder–decoder could not achieve the desired performance when N_{enc} was reduced to lower than 32.

4. Results and discussion

4.1. Interpolation prediction of trained encoder–decoder

Several examples of post-impact image sequences generated by the trained encoder–decoder from the pre-impact image sequences with the H values reserved for testing are shown alongside the pre-impact inputs and the actual output image sequences in figure 8. The heat maps shown next to the actual image sequences show the differences between the respective generated and actual image sequences. As shown, the generated and actual image sequences are generally in good agreement. For non-splashing drops in particular, the difference is represented only by a thin line along the contour of the drop.

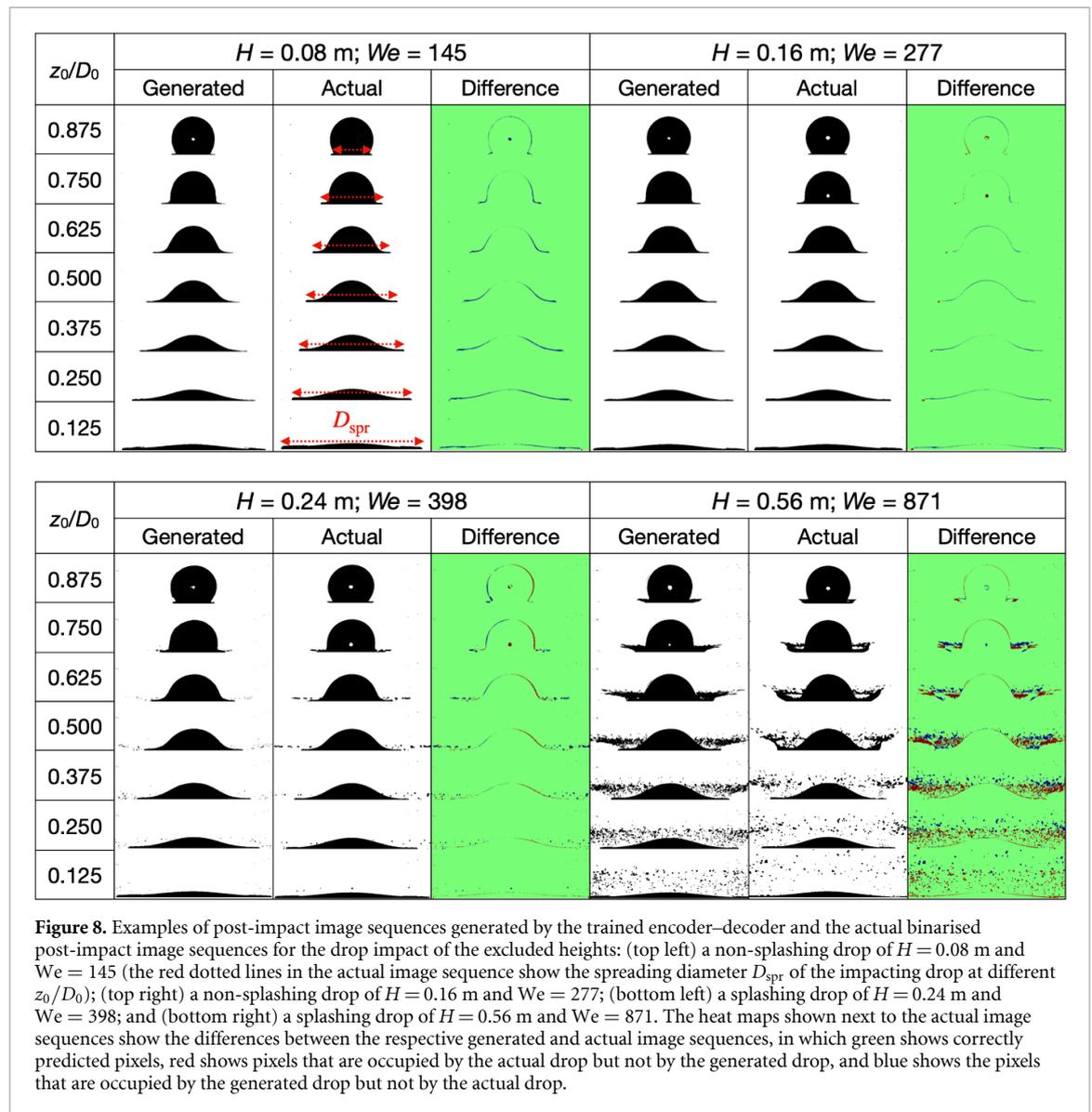
For splashing drops of $H = 0.56$ m, careful examination of the generated and actual image sequences reveals a difference in the ejection angle of the secondary droplets. This can be seen from the example shown in figure 8 (bottom right) in the splashing drop of $H = 0.56$ m and $We = 871$. Especially for $0.500 \leq z_0/D_0 \leq 0.750$ (the second to fourth frames), the secondary droplets in the generated image sequences are ejected outwards, forming a shape like a plate, while those in the actual image sequences are ejected upwards, forming a shape like a bowl. Nevertheless, the results are satisfying because there is a clear difference in the distribution and number of ejected secondary droplets between the image sequences generated using different H values. This can be seen when comparing the splashing drops of $H = 0.24$ m and $We = 398$ (bottom left) and $H = 0.56$ m and $We = 871$ (bottom right). These findings demonstrate the ability of the trained encoder–decoder to generate image sequences of splashing drops of different morphologies resulting from different physical parameters.

4.2. Spreading diameter of impacting drop

For quantitative evaluation of the generated post-impact image sequences, the spreading diameter D_{spr} of the impacting drop in each frame (z_0/D_0) of the image sequence generated by the trained encoder–decoder for each set of testing data was measured according to the definition shown in figure 8 and normalised by D_0 .

Figure 9(a) shows the mean normalised spreading diameters D_{spr}/D_0 of the impacting drops averaged among each frame of the generated and actual image sequences of the interpolation impact heights H , with the error bars showing the standard deviations. In figure 9(a), for each interpolation H value, the mean D_{spr}/D_0 values of the generated and actual image sequences of the impacting drop for each z_0/D_0 overlap, indicating excellent agreement between the two. The prediction error is quantified by computing the percentage difference between the D_{spr} values of the impacting drop in each frame of the generated and actual image sequences.

Plots of the prediction error for each z_0/D_0 are shown in figure 9(b). Other than for $z_0/D_0 = 0.875$, all of the prediction errors are low: between -15% and 15% . The prediction error for $z_0/D_0 = 0.875$ is relatively



high because the lamella is not fully developed; for some non-splashing drops, a lamella is not ejected, whereas for some splashing drops, it is lifted up. These phenomena were considered by Riboux and Gordillo [6, 7], who stated that the ejection time of the lamella scales with We when the Ohnesorge number Oh is sufficiently small, and they attributed splashing to the lift force acting on the lamella.

To further analyse the behaviour of D_{spr} , the D_{spr}/D_0 values of the generated and actual image sequences were averaged among each interpolation H value and plotted against z_0/D_0 , as shown in figure 9(c), in which the error bars show the standard deviations. The distribution shows that there is a negative correlation between D_{spr}/D_0 and z_0/D_0 . This is because as time elapses, the central height of the drop z_0 decreases as the lamella continues to spread, thus D_{spr} increases. Similar to figure 9(a), for all values of z_0/D_0 , the mean D_{spr}/D_0 values of the generated and actual image sequences of the impacting drop for each interpolation H value overlap, indicating excellent agreement between the two.

Plots of the prediction error for each H value are shown in figure 9(d). Other than for $H = 0.08$ and 0.32 m , all of the prediction errors are again low: between -20% and 20% . The prediction error of $H = 0.08 \text{ m}$ is relatively high because the lamella is ejected later due to the low We value. Thus, the standard deviation of D_{spr} is higher, especially for $z_0/D_0 = 0.875$.

4.3. Splashing/non-splashing prediction

The accuracy of splashing/non-splashing prediction of the trained encoder–decoder was evaluated by inspecting the impacting drop in each frame (z_0/D_0) of the generated image sequences. To determine whether an impacting drop is splashing, the main criteria are the presence of ejected secondary droplets and

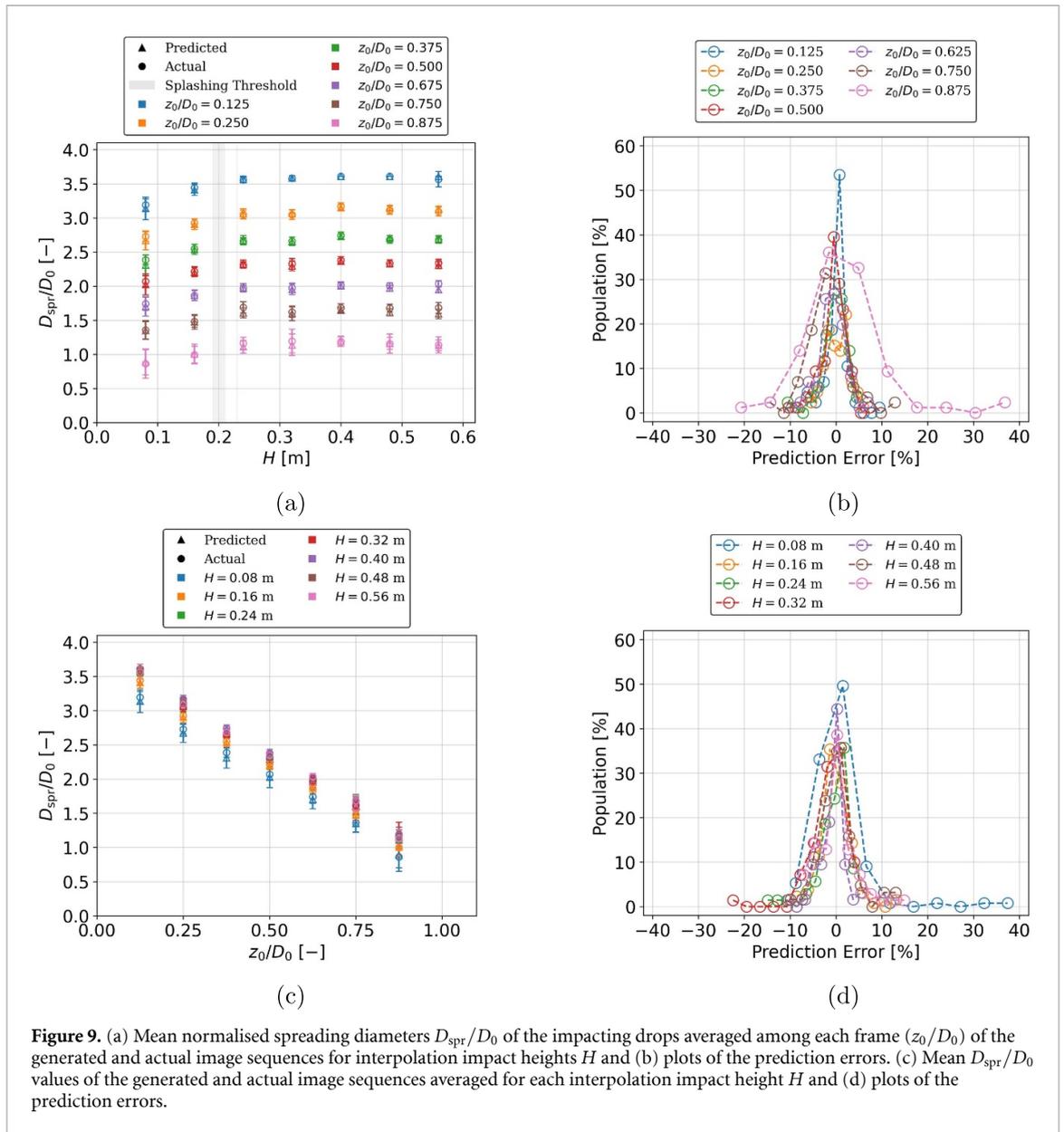
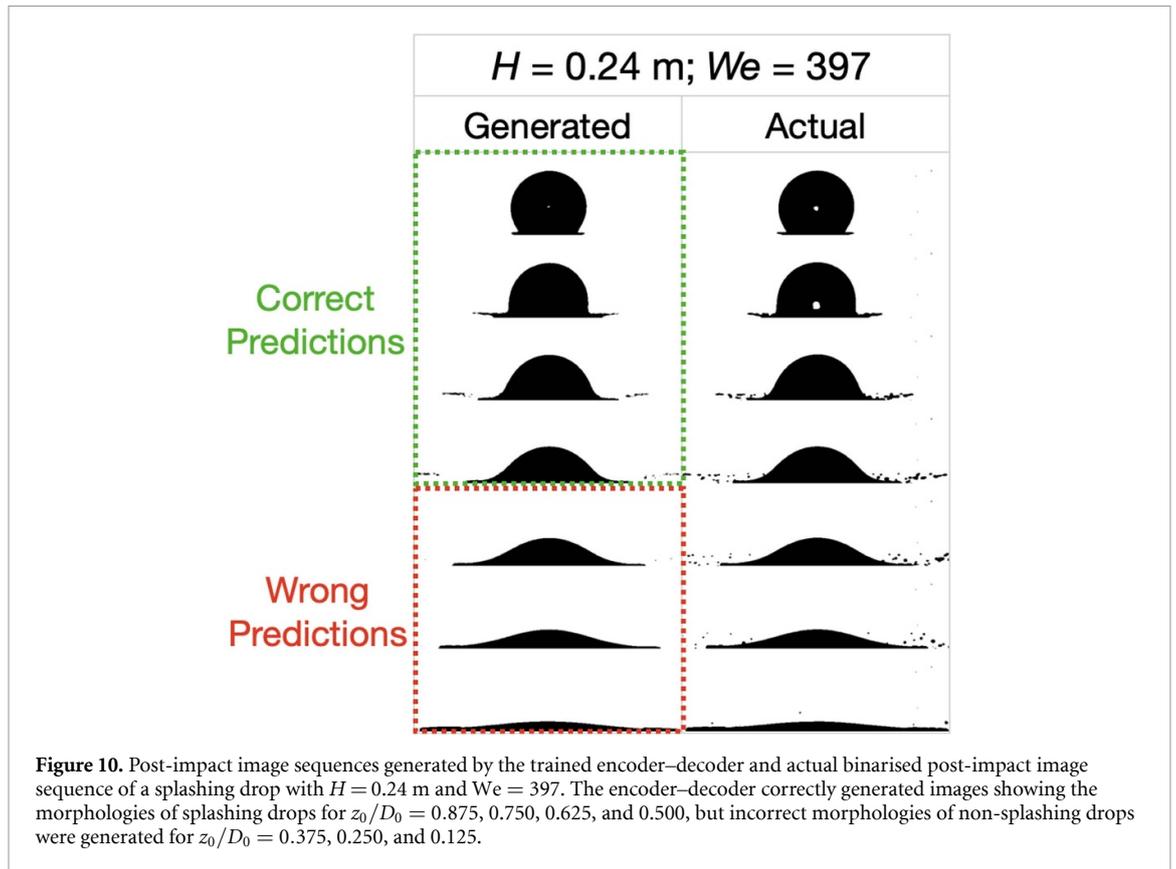


Figure 9. (a) Mean normalised spreading diameters D_{spr}/D_0 of the impacting drops averaged among each frame (z_0/D_0) of the generated and actual image sequences for interpolation impact heights H and (b) plots of the prediction errors. (c) Mean D_{spr}/D_0 values of the generated and actual image sequences averaged for each interpolation impact height H and (d) plots of the prediction errors.

the behaviour of the lamella. Although a splashing drop is mainly characterised by the ejection of secondary droplets, for some splashing drops, there are very few detached secondary droplets at the early stage of the impact when $z_0/D_0 = 0.75$ (see bottom left of figure 8). Nevertheless, the splashing of these drops can still be identified by the lifted lamella [7].

It is important to point out that the splashing/non-splashing prediction is not necessarily the same for all of the frames in a generated image sequence. One example is the image sequence generated for the splashing drop of $H = 0.24$ m and $We = 397$ that is shown alongside the actual image sequence in figure 10. In this figure, the encoder–decoder correctly generated images that show the morphologies of splashing drops for $z_0/D_0 = 0.875, 0.750, 0.625,$ and 0.500 , but it generated images showing incorrect morphologies of non-splashing drops for $z_0/D_0 = 0.375, 0.250,$ and 0.125 . Thus, instead of evaluating the splashing/non-splashing prediction of a generated image sequence as a whole, it is necessary to perform a frame-by-frame or z_0/D_0 -by- z_0/D_0 evaluation.

The accuracy of splashing/non-splashing prediction for each value of z_0/D_0 of the generated image sequences is shown by the bar chart in figure 11(a). A high prediction accuracy was achieved, with an overall accuracy level greater than 79%. For all z_0/D_0 values except 0.125, the splashing prediction accuracy was higher than the non-splashing prediction accuracy. This is because when $z_0/D_0 = 0.125$, the ejected secondary droplets are widely scattered away from the impacting drop, and it is therefore difficult for the trained encoder–decoder to capture and reproduce this. However, when $z_0/D_0 \geq 0.625$, the secondary



droplets are only just ejected from the impacting drop and are still accumulating around the lamella; thus, the splashing prediction accuracies are higher than 90%.

The accuracy of splashing/non-splashing prediction for each interpolation impact height H was also evaluated, and the results are shown in the bar chart in figure 11(b). Note that the accuracy was computed frame-by-frame for every generated image sequence. For example, for the generated image sequences shown in figure 10, four out of the seven frames were correctly predicted, hence the accuracy was 57%. Since there is a splashing transition within $0.20 \text{ m} \leq H \leq 0.22 \text{ m}$, there are only non-splashing and splashing prediction accuracies for $H \leq 0.16 \text{ m}$ and $H \geq 0.24 \text{ m}$, respectively. Similar to figure 11(a), the splashing prediction accuracy is higher than the non-splashing prediction accuracy, except for at $H = 0.24 \text{ m}$. The prediction accuracy is very low for $H = 0.24 \text{ m}$ because this is the lowest of the splashing H values. In other words, the impact velocity is the lowest for $H = 0.24 \text{ m}$, and this means that splashing features such as the number of ejected secondary droplets are least prevalent. Nevertheless, the splashing prediction accuracy increases with H and eventually reaches a perfect 100% for $H \geq 0.48 \text{ m}$.

A more detailed analysis was performed, and the prediction accuracy for each value of z_0/D_0 at each interpolation H value is shown by the heat map in figure 11(c). Similar to figures 11(a) and (b), it can be seen that the splashing prediction accuracy ($H \geq 0.24 \text{ m}$) is higher than the non-splashing prediction accuracy ($H \leq 0.16 \text{ m}$), except for $H = 0.24 \text{ m}$, which has the lowest prediction accuracy. Additionally, the prediction accuracy is a perfect 100% for each z_0/D_0 for interpolation $H \geq 0.48 \text{ m}$. For splashing H values, the prediction accuracy tends to decrease with z_0/D_0 . This shows that the trained encoder–decoder tends to predict a splashing drop for high z_0/D_0 values and a non-splashing drop for low z_0/D_0 values.

4.4. Analysis on the prediction process

Analysis was performed to understand the prediction process by the encoder–decoder, specifically, the image features of the pre-impact image sequences that the trained encoder identifies and how these image features affect the generation of the post-impact image sequences by the trained decoder. However, there are too many parameters involved in the prediction process, the important parameters had to be first identified by extracting the important encoded elements.

The encoded elements, which are shown by the green nodes in figure 7, play important roles as they connect the encoder and the decoder. The important elements were extracted by analysing the $s_{h,i}$ computed

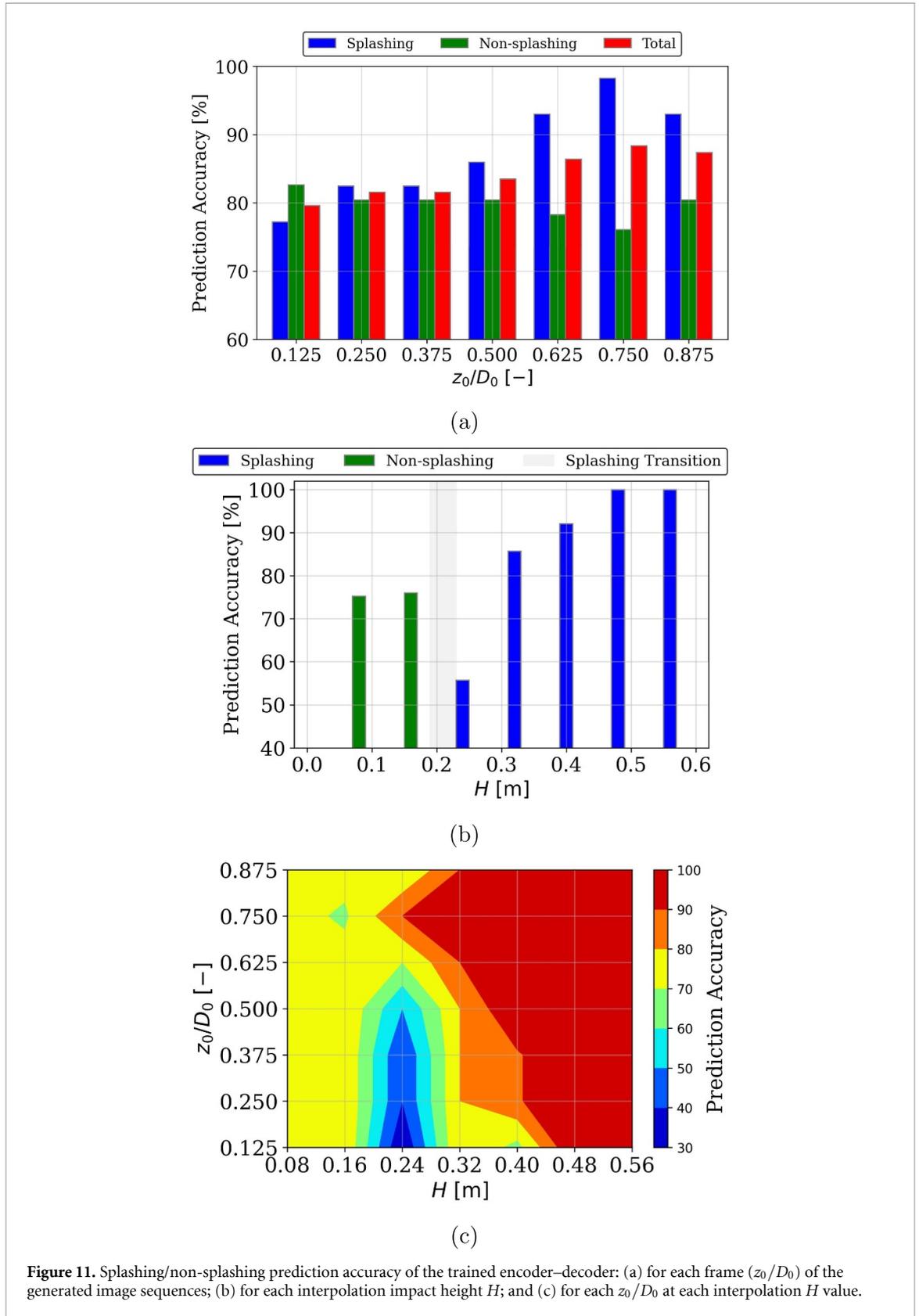
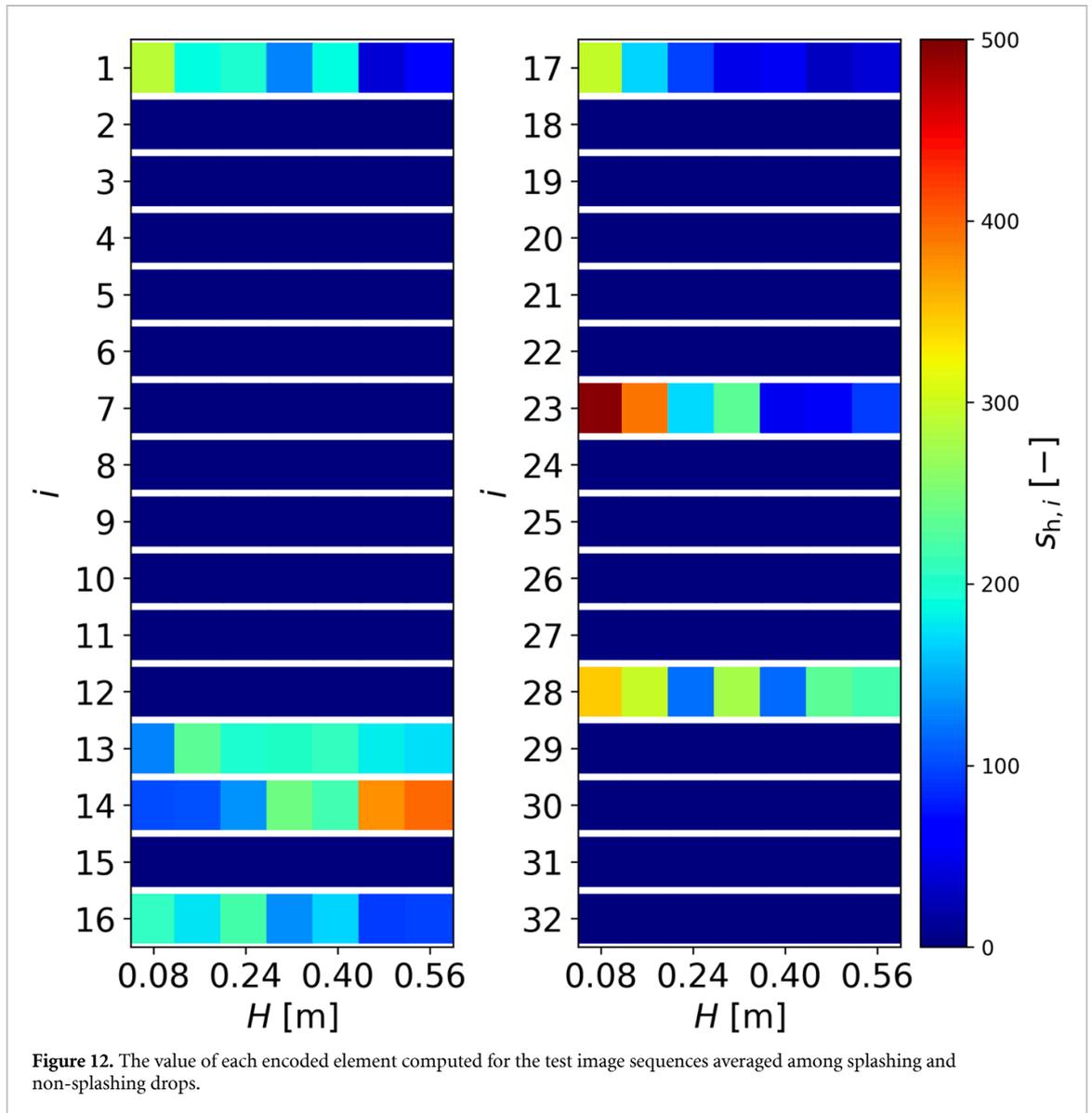


Figure 11. Splashing/non-splashing prediction accuracy of the trained encoder–decoder: (a) for each frame (z_0/D_0) of the generated image sequences; (b) for each interpolation impact height H ; and (c) for each z_0/D_0 at each interpolation H value.

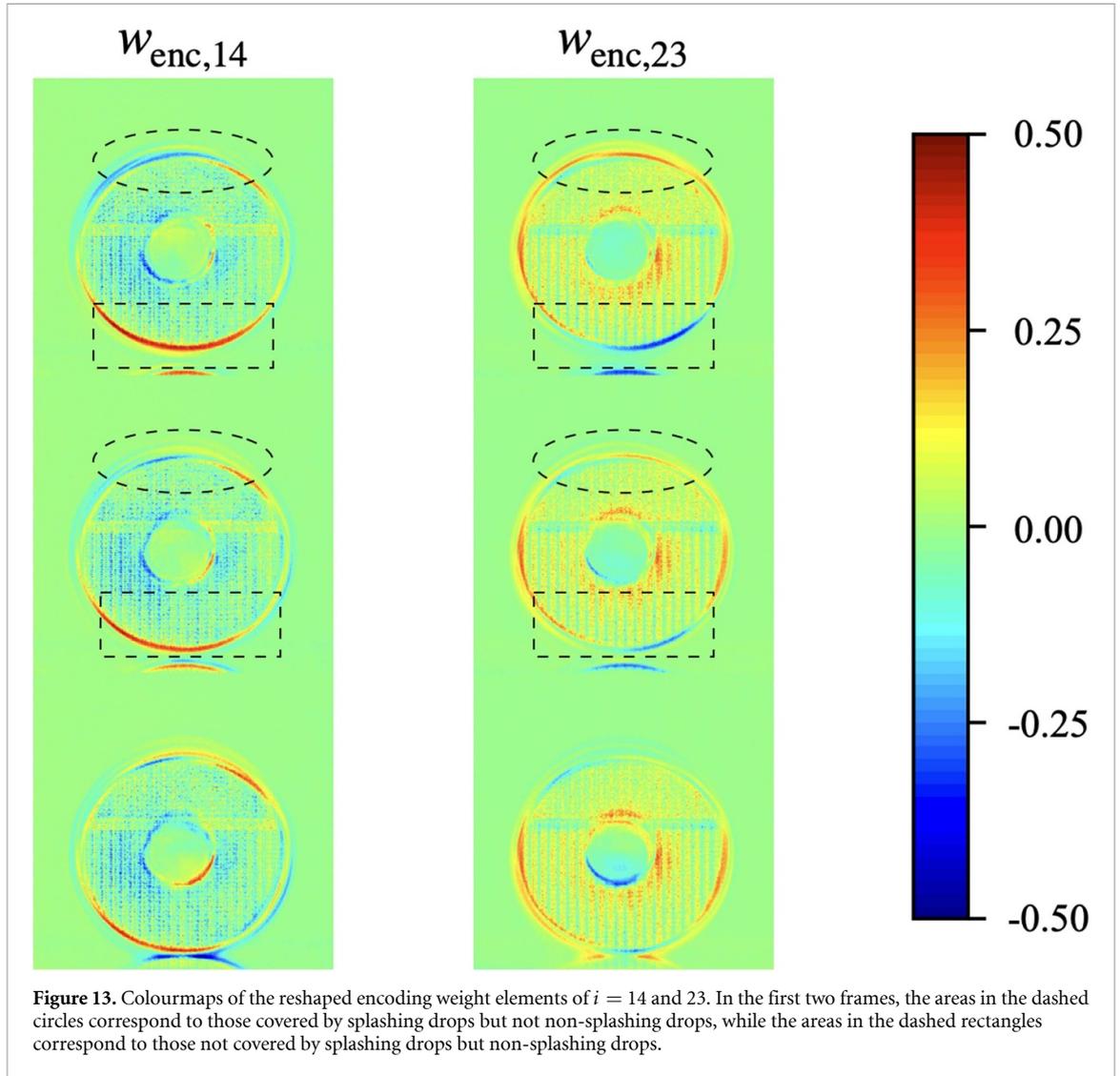
for the test image sequences. Here, $i = 1, \dots, N_{\text{enc}}$, where N_{enc} was set to 32. Based on the pre-impact image sequences, the encoder computes the values of $s_{h,i}$. The values of $s_{h,i}$ were averaged among the test image sequences of the same H and plotted in figure 12. The elements with the value $s_h \approx 0$ were inactive as they did not affect the computation of $s_{\text{out},i}$, i.e. the values of the intensity in the post-impact image sequence that form the drop morphologies. On the other hand, the active elements are $i = 1, 13, 14, 16, 17, 23,$ and 28 .



These active elements are important because their values were used by the decoder to compute the values of $s_{out,i}$.

Instead of all the encoded elements, the explanation of the prediction process of the trained encoder-decoder here focuses on the analysis of the elements of $i = 14$ and 23 . These two elements are good examples because they have the highest range of $s_{h,i}$. Besides, their relationships with H are opposite to one another. As H increases from 0.08 to 0.56 m, $s_{h,14}$ increases from ≈ 100 to ≈ 500 while $s_{h,23}$ decreases from ≈ 500 to ≈ 100 . These indicate that a splashing drop has a high value of $s_{h,14}$ and a low value of $s_{h,23}$ while a non-splashing drop has a low value of $s_{h,14}$ and a high value of $s_{h,23}$.

To understand the image features extracted by the encoder from the pre-impact image sequences, the elements of the encoding weight matrix \mathbf{W}_{enc} of $i = 14$ and 23 were visualised. For this purpose, the encoding weight elements $w_{enc,14}$ and $w_{enc,23}$ were reshaped into the shape of a pre-impact image sequence and presented in figure 13 with the blue-green-red (BGR) scale from -0.50 to 0.50 . Since splashing drops have a higher U_0 than non-splashing drops, the splashing drops cover higher positions than the non-splashing drops in the first two frames of a pre-impact image sequence. The respective positions are shown in figure 13, where the dashed circles show the area covered by splashing drops but not by non-splashing drops, while the dashed rectangles show the area not covered by splashing drops but by non-splashing drops. Observation on these areas shows that in the colourmap of $w_{enc,14}$, there are more negative values (blue) in the dashed circles, while there are more positive values (red) in the dashed rectangles. Thus, the negative values remain for a non-splashing drop leading to a lower value of $s_{h,14}$, while the positive values remain for a splashing drop leading to a higher value of $s_{h,14}$. On the other hand, in the

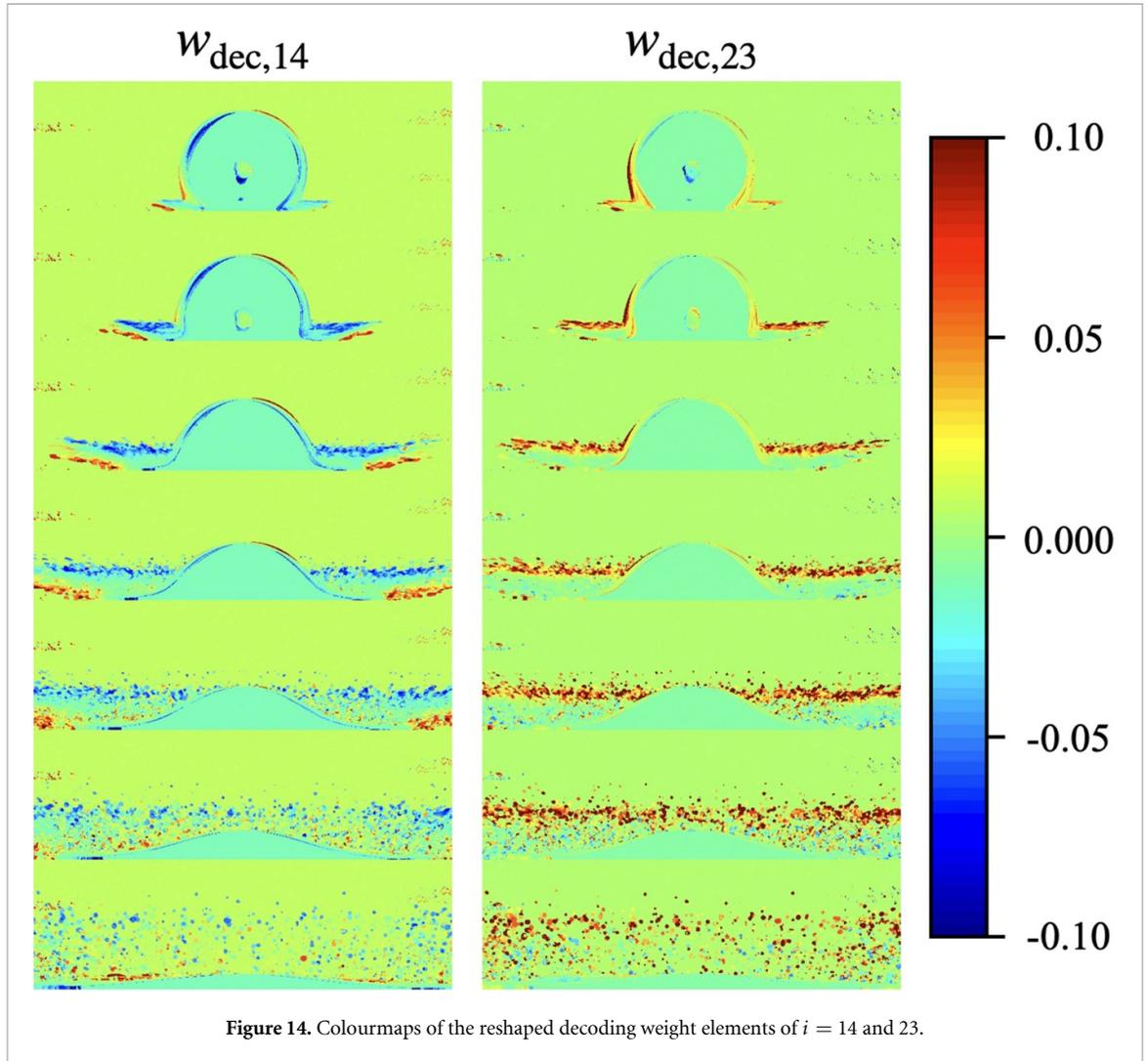


colormap of $w_{enc,23}$, there are more positive values (red) in the dashed circles, while there are more negative values (blue) in the dashed rectangles. Thus, the negative values remain for a non-splashing drop leading to a lower value of $s_{h,23}$, while the positive values remain for a splashing drop leading to a higher value of $s_{h,23}$.

To understand how the decoder computes the intensity values of each pixel in the post-impact image sequences to predict the drop morphologies, the elements of the decoding weight matrix \mathbf{W}_{dec} of $i = 14$ and 23 were visualised. For this purpose, the decoding weight elements $w_{dec,14}$ and $w_{dec,23}$ were reshaped into the shape of a post-impact image sequence and presented as colormaps in figure 14 with the BGR scale from -0.10 to 0.10 . Analysis was then performed on $w_{dec,14}$ and $w_{dec,23}$.

Since the value of $s_{h,14}$ is high for splashing drops but low for non-splashing drops, there are more negative values (blue) in the colormaps of $w_{dec,14}$. For splashing drops, since $s_{h,14}$ is high, these negative values remain and become zero after being activated by the sigmoid function (equation (6)). These zero values form the morphology of a drop. On the other hand, for non-splashing drops, since $s_{h,14}$ is low, these negative values do not remain but become non-zero values after being activated by the sigmoid function. These non-zero values do not form the morphology of a drop. Note that there are some positive values under the distribution of negative values that forms the morphology of a lamella. This can be related to the higher ejection angle of the lamella of a corona splash, which occurs at a higher velocity than a prompt splash.

A similar but opposite explanation can be made for the element of $i = 23$. Since the value of $s_{h,23}$ is low for splashing drops but high for non-splashing drops, there are more positive values (red) in the colormaps of $w_{dec,23}$. For splashing drops, since $s_{h,23}$ is low, these positive values do not remain but become zero values that form the morphology of a drop. On the other hand, for non-splashing drops, since $s_{h,14}$ is high, these positive values remain and become non-zero values, thus not forming the morphology of a drop.



5. Conclusion

In this study, an encoder–decoder model was trained to generate videos in the form of image sequences that can accurately represent the actual morphological evolutions of splashing and non-splashing drops during their impact on a solid surface under different impact velocities.

Interpolation prediction was performed by reserving the image sequences of drops of certain interpolation impact heights H for testing. The image sequences generated for the interpolation H values showed good agreement with the actual image sequences. Notably, among splashing drops of different H , there was a clear difference in morphology in terms of the distribution and number of ejected secondary droplets.

The spreading diameters D_{spr} of the impacting drops in the generated image sequences showed excellent agreement with those in the actual image sequences. Averaging among the frames of each normalised central height of the impacting drops z_0/D_0 , the prediction error was $\pm 15\%$ for $z_0/D_0 \leq 0.750$. Averaging among H , the prediction error was $\pm 20\%$ for H values other than 0.08 and 0.32 m. The prediction error can be related to the development of the lamella, and the error is higher at the beginning of the lamella ejection.

The overall accuracy of splashing/non-splashing prediction of the generated image sequences was greater than 79%. The trained encoder–decoder tends to generate images of splashing drops for frames of higher z_0/D_0 , while for frames of lower z_0/D_0 , it tends to generate images of non-splashing drops. This is because more ejected secondary droplets accumulate around an impacting splashing drop when z_0/D_0 is high; thus, those splashing features can be easily captured by the trained encoder–decoder. Conversely, when z_0/D_0 is low, the ejected droplets are widely scattered away from the impacting drop, and it is thus difficult for this to be captured by the trained encoder–decoder.

The findings of this study demonstrate the ability of the trained encoder–decoder to generate image sequences that can accurately represent the morphologies of splashing and non-splashing drops for

intermediate data points. This approach provides a faster and cheaper alternative to experiments and numerical simulations. The ability of the trained encoder–decoder could be developed for researching other phenomena, especially multiphase flows.

Data availability statement

The data cannot be made publicly available upon publication because no suitable repository exists for hosting data in this field of study. The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Jingzu Yee  <https://orcid.org/0000-0002-0549-165X>

Yoshiyuki Tagawa(田川義之)  <https://orcid.org/0000-0002-0049-1984>

References

- [1] Hu S, Zhu W, Yang W and Li M 2022 *npj Flex. Electron.* **6** 64
- [2] Sun T P, Álvarez-Novoa F, Andrade K, Gutiérrez P, Gordillo L and Cheng X 2022 *Nat. Commun.* **13** 1703
- [3] Modak C D, Kumar A, Tripathy A and Sen P 2020 *Nat. Commun.* **11** 4327
- [4] Breitenbach J, Roisman I V and Tropea C 2018 *Exp. Fluids* **59** 55
- [5] De Ruiter J, Lagraauw R, Van Den Ende D and Mugele F 2015 *Nat. Phys.* **11** 48–53
- [6] Riboux G and Gordillo J M 2017 *Phys. Rev. E* **96** 013105
- [7] Riboux G and Gordillo J M 2014 *Phys. Rev. Lett.* **113** 024507
- [8] Yokoyama Y, Tanaka A and Tagawa Y 2022 *Forensic Sci. Int.* **331** 111138
- [9] Hatakenaka R, Breitenbach J, Roisman I V, Tropea C and Tagawa Y 2019 *Int. J. Heat Mass Transfer* **145** 118729
- [10] Rioboo R, Tropea C and Marengo M 2001 *At. Sprays* **11** 155–65
- [11] Liu X, Zhang X and Min J 2019 *Phys. Fluids* **31** 092102
- [12] Gordillo J M, Riboux G and Quintero E S 2019 *J. Fluid Mech.* **866** 298–315
- [13] Lin S, Zhao B, Zou S, Guo J, Wei Z and Chen L 2018 *J. Colloid Interface Sci.* **516** 86–97
- [14] Huang H M and Chen X P 2018 *Phys. Fluids* **30** 022106
- [15] Clanet C, Béguin C, Richard D and Quéré D 2004 *J. Fluid Mech.* **517** 199–208
- [16] Fernández-Raga M, Palencia C, Keesstra S, Jordán A, Fraile R, Angulo-Martínez M and Cerdà A 2017 *Earth-Sci. Rev.* **171** 463–77
- [17] Waite H, Whitelaw-Weckert M and Torley P 2015 *N. Z. J. Crop Hortic. Sci.* **43** 144–61
- [18] Lohse D 2022 *Annu. Rev. Fluid Mech.* **54** 349–82
- [19] Markt J D, Raessi M, Lee S Y and Zhu X 2021 *Phys. Fluids* **33** 102120
- [20] Philipp J, Lagrée P Y and Antkowiak A 2016 *J. Fluid Mech.* **795** 96–135
- [21] Lagubeau G, Fontelos M A, Josserand C, Maurel A, Pagneux V and Petitjeans P 2012 *J. Fluid Mech.* **713** 50–60
- [22] Eggers J, Fontelos M A, Josserand C and Zaleski S 2010 *Phys. Fluids* **22** 062101
- [23] Vo Q and Tran T 2021 *Nat. Commun.* **12** 4289
- [24] Liu Q, Lo J H Y, Li Y, Liu Y, Zhao J and Xu L 2021 *Nat. Commun.* **12** 3068
- [25] de Ruiter J, Soto D and Varanasi K K 2018 *Nat. Phys.* **14** 35–39
- [26] Gordillo L, Sun T P and Cheng X 2018 *J. Fluid Mech.* **840** 190–214
- [27] Josserand C and Thoroddsen S T 2016 *Annu. Rev. Fluid Mech.* **48** 365–91
- [28] Yarin A L 2006 *Annu. Rev. Fluid Mech.* **38** 159–92
- [29] Usawa M, Fujita Y, Tagawa Y, Riboux G and Gordillo J M 2021 *Phys. Rev. Fluids* **6** 023605
- [30] Zhang H, Zhang X, Yi X, Du Y, He F, Niu F and Hao P 2022 *Phys. Fluids* **34** 022111
- [31] Zhang H, Zhang X, Yi X, He F, Niu F and Hao P 2021 *Phys. Fluids* **33** 052103
- [32] Jalali B, Zhou Y, Kadambi A and Roychowdhury V 2022 *Mach. Learn.: Sci. Technol.* **3** 041001
- [33] Touranakou M, Chernyavskaya N, Duarte J, Gunopulos D, Kansal R, Orzari B, Pierini M, Tomei T and Vlimant J R 2022 *Mach. Learn.: Sci. Technol.* **3** 035003
- [34] Vlachas P R, Arampatzis G, Uhler C and Koumoutsakos P 2022 *Nat. Mach. Intell.* **4** 359–66
- [35] Knijff L and Zhang C 2021 *Mach. Learn.: Sci. Technol.* **2** 03LT03
- [36] Thavarajah R, Zhai X, Ma Z and Castineira D 2021 *Mach. Learn.: Sci. Technol.* **2** 025022
- [37] Fukami K, Maulik R, Ramachandra N, Fukagata K and Taira K 2021 *Nat. Mach. Intell.* **3** 945–51
- [38] Novati G, de Laroussilhe H L and Koumoutsakos P 2021 *Nat. Mach. Intell.* **3** 87–96
- [39] Brunton S L, Noack B R and Koumoutsakos P 2020 *Annu. Rev. Fluid Mech.* **52** 477–508
- [40] Yancheshme A A, Enayati S, Kashcooli Y, Jafari R, Ezzaidi H and Momen G 2022 *Exp. Therm. Fluid Sci.* **139** 110743
- [41] Tembely M, Vadillo D C, Dolatabadi A and Soucemarianadin A 2022 *Processes* **10** 1141
- [42] Yoon I, Chergui J, Juric D and Shin S 2022 *Phys. Fluids* **34** 102109
- [43] Pierzyna M, Burzynski D A, Bansmer S E and Semaan R 2021 *Phys. Fluids* **33** 123317
- [44] Dickerson A K, Alam M E, Buckelew J, Boyum N and Turgut D 2022 *Phys. Fluids* **34** 102112
- [45] Yee J, Yamanaka A and Tagawa Y 2022 *Phys. Fluids* **34** 013317
- [46] Kingma D P and Welling M 2014 Auto-encoding variational Bayes *2nd Int. Conf. on Learning Representation (ICLR 2014)*
- [47] Lu L, Jin P, Pang G, Zhang Z and Karniadakis G E 2021 *Nat. Mach. Intell.* **3** 218–29
- [48] Krenn M, Häse F, Nigam A, Friederich P and Aspuru-Guzik A 2020 *Mach. Learn.: Sci. Technol.* **1** 045024
- [49] Käming N, Dawid A, Kottmann K, Lewenstein M, Sengstock K, Dauphin A and Weitenberg C 2021 *Mach. Learn.: Sci. Technol.* **2** 035037
- [50] Badrinarayanan V, Kendall A and Cipolla R 2017 *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 2481–95
- [51] Khan S D, Alarabi L and Basalamah S 2023 *Arab. J. Sci. Eng.* **48** 1273–84

- [52] Khan S D, Alarabi L and Basalamah S 2020 *Computers* **9** 99
- [53] Gonzalez R C, Eddins S L and Woods R E 2004 *Digital Image Publishing Using Matlab* (New Jersey: Pearson Prentice Hall)
- [54] Otsu N 1979 *IEEE Trans. Syst. Man Cybern.* **9** 62–66
- [55] Buades A, Coll B and Morel J M 2005 A non-local algorithm for image denoising 2005 *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05)* vol 2 (IEEE) pp 60–65
- [56] Atherton T J and Kerbyson D J 1999 *Image Vis. Comput.* **17** 795–803
- [57] Yuen H K, Princen J, Illingworth J and Kittler J 1990 *Image Vision Comput.* **8** 71–77
- [58] Canny J 1986 *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8** 679–98
- [59] Hirt C W and Nichols B D 1981 *J. Comput. Phys.* **39** 201–25
- [60] Kluyver T et al 2016 Jupyter notebooks—a publishing format for reproducible computational workflows *Positioning and Power in Academic Publishing: Players, Agents and Agendas* ed F Loizides and B Schmidt (Amsterdam: IOS Press BV) pp 87–90
- [61] Abadi M et al 2016 Tensorflow: a system for large-scale machine learning *12th USENIX Symp. on Operating Systems Design and Implementation (OSDI '16)* pp 265–83
- [62] Kingma D P and Ba J 2014 (arXiv:[1412.6980](https://arxiv.org/abs/1412.6980))