Deep learning based liquid level extraction from video observations of gas–liquid flows

Marc Olbrich [a,b,*], Leili Riazy [c], Tobias Kretz [a], Terri Leonard [d], Dennis S. van Putten [e], Markus Bär [a], Kilian Oberleithner [b], Sonja Schmelter [a]

[a] *Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2-12, 10587 Berlin, Germany*
[b] *Institute of Fluid Dynamics and Technical Acoustics, Technische Universität Berlin, Müller-Breslau-Straße 8, 10623 Berlin, Germany*
[c] *Cardiovascular Magnetic Resonance, Experimental and Clinical Research Center, Charité Campus Buch, Lindenberger Weg 80, 13125 Berlin, Germany*
[d] *TÜV SÜD National Engineering Laboratory (NEL), East Kilbride, Glasgow G75 0QF, United Kingdom*
[e] *DNV, Energieweg 17, 9743 AN Groningen, The Netherlands*

ARTICLE INFO

ABSTRACT

The slug flow pattern is one of the most common gas–liquid flow patterns in multiphase transportation pipelines, particularly in the oil and gas industry. This flow pattern can cause severe problems for industrial processes. Hence, a detailed description of the spatial distribution of the different phases in the pipe is needed for automated process control and calibration of predictive models. In this paper, a deep-learning based image processing technique is presented that extracts the gas–liquid interface from video observations of multiphase flows in horizontal pipes. The supervised deep learning model consists of a convolutional neural network, which was trained and tested with video data from slug flow experiments. The consistency of the hand-labelled data and the predictions of the trained model have been evaluated in an inter-observer reliability test. The model was further tested with other data sets, which also included recordings of a different flow pattern. It is shown that the presented method provides accurate and reliable predictions of the gas–liquid interface for slug flow as well as for other separate flow patterns. Moreover, it is demonstrated how flow characteristics can be obtained from the results of the deep-learning based image processing technique.

1. Introduction

Multiphase flow phenomena are often encountered in different sectors of the energy industry, particularly in the oil and gas production, where the two phases of liquid and gas are flowing simultaneously through transportation pipelines (Lin et al., 2020). Field measurements of these flows have a high degree of uncertainty, reaching up to 20% (Elliott et al., 2021). Based on the operating conditions, different flow patterns can form, which describe the spatial distribution of the two phases in the pipe (Hanratty, 2013). One of the most common flow patterns in multiphase transportation pipelines is the slug flow pattern (Al-Kayiem et al., 2017).

Slug flow is characterized by a continuous liquid phase with coherent blocks of aerated liquid, which are separated by volumes of gas, see Fig. 1, left. These aerated blocks of liquid are called slugs. They are moving downstream the pipe on top of a slowly flowing liquid layer at approximately the same velocity as the gas (Hanratty, 2013; Taitel and Dukler, 1977; Al-Safran, 2009). The slug flow pattern can cause severe problems in industrial operations. The pressure drop due

to slug flow can be an order of magnitude higher compared to other gas–liquid flow patterns (Taitel and Dukler, 1977), such as stratified or wavy flow, where the two phases are separated by a smooth or wavy interface (Hanratty, 2013). Furthermore, the intermittent sequence of liquid slugs, which can grow to large structures, induces vibrations and stresses when impacting on surfaces of the piping (Hanratty, 2013). These problems can also affect the multiphase flow measurement devices significantly. Thus, a detailed characterization of the flow pattern is required and had already been the subject of many investigations, see, e.g., Pedersen et al. (2017), Olbrich et al. (2020) and Drury et al. (2019). Parameters that are typically used for a characterization of this flow pattern are temporal and spatial scales of the liquid slugs as well as their translational velocities. Spatial scales are for example the slug body length L_b and the slug unit length L_u, i.e. the distance between the slug front and slug rear or between two consecutive slug fronts, respectively. Their corresponding time scales, the slug unit time T_u and the slug body time T_b, are defined as the time that a slug body or a slug unit needs to pass by at a fixed position. Another often used parameter
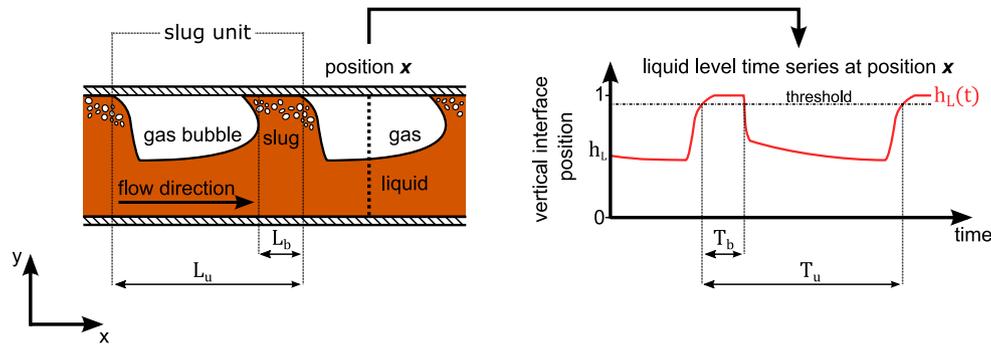

Fig. 1. Illustration of slug flow in a horizontal pipe (left side) with a slug unit, the slug unit length L_u, and the slug body length L_b, as well as the liquid level time series $h_L(t)$ at a fixed position x (right side) with the slug unit time T_u and the slug body time T_b for a given threshold.

is the mean slug frequency, defined as

$$\bar{f}_s = \frac{1}{\bar{T}_u} \quad \text{with} \quad \bar{T}_u = \frac{1}{N_s} \sum_{i=1}^{N_s} T_{u_i}, \tag{1}$$

where N_s denotes the number of slug units in the considered time interval, see Al-Kayiem et al. (2017), Baba et al. (2018), Dukler and Fabre (1994) and Olbrich et al. (2021a). The length and time scales of slug flow are illustrated in Fig. 1. These characteristics are often measured with non-intrusive imaging techniques, such as videometric approaches, where the flow is observed with a high-speed camera through a transparent pipe segment (do Amaral et al., 2013). The flow parameters provided by these type of measurement techniques were used for example to identify the flow pattern (Baghernejad et al., 2019), to verify and investigate flow pattern maps (Crawford, 2018), to investigate shapes of slugs and bubbles for specific operating conditions (do Amaral et al., 2013), to validate numerical simulations (Olbrich et al., 2018), as well as to investigate the effects of slug frequency on induced pipe stresses and develop predictive models and correlations (Mohmmed et al., 2019).

One way to obtain these characteristic parameters is to consider the time series of the vertical position of the liquid–gas interface at a fixed x-position in the pipe, as illustrated on the right side of Fig. 1. This non-dimensional parameter has a range of $[0, 1]$ with respect to the inner pipe diameter D and is hereinafter referred to as the *liquid level time series* at a fixed position x, denoted by $h_L(t)$ (Olbrich et al., 2021a; Schmelter et al., 2021a). It reveals the dynamics of the spatial distribution of the two phases in the pipe and can therefore reliably indicate slugs, waves or other liquid structures, similarly to the hold-up parameter, see Olbrich et al. (2021b). Typically, for analyses of hold-up or liquid level time series, the conventional length and time scales of slug flow are calculated by simple thresholding procedures, see Zhao et al. (2015), Baba et al. (2018), Schmelter et al. (2021a) and Olbrich et al. (2020, 2021b) and Fig. 1.

In this paper, the liquid level time series are derived from high-speed video recordings of gas–liquid flows observed from the side through a transparent pipe section. These flows are two-phase gas–oil and gas–water flows, as well as three-phase gas–oil–water flows, where the liquid phase is a homogeneous mixture of oil and water. For this, a deep learning based image processing technique has been developed. In an earlier work, the time series have been extracted from the video data using a fixed sequence of image filters, see Olbrich et al. (2018). This was similar to other approaches in the field, see e.g., do Amaral et al. (2013) and provided reasonable results for the video data with (exactly) the same conditions it was developed for, such as the colours of the fluids and the background as well as lighting and reflections. However, changes in these conditions as well as noisy data led to incorrect liquid level estimations, and individual adaptations were needed. In contrast to this, deep learning models have the potential to overcome such difficulties and to provide reliable and more versatile image processing techniques.

Deep learning describes a family of learning algorithms in the field of machine learning and artificial intelligence (Emmert-Streib et al., 2020). It is used to learn complex and robust prediction models, e.g., multi-layer neural networks with many hidden units, directly from the data without the need of carefully engineering suitable features (Emmert-Streib et al., 2020; LeCun et al., 2015). Deep learning has a wide range of applications in science, business, and technology, e.g., image or speech recognition, see LeCun et al. (2015), Krizhevsky et al. (2017), Ronneberger et al. (2015) and Graves et al. (2013). However, in the field of multiphase flows, deep learning has only rarely been applied. In the following, we give a short summary on such applications.

For the numerical simulations of multiphase flows, deep learning models were trained for example to approximate the governing equations, estimate simulation errors, predict flow parameters, as well as closure coefficients, see e.g., Wang and Lin (2020), Bao et al. (2020) and Ma et al. (2015). Other applications of deep learning models are for example the correction or prediction of certain parameters for multiphase flow measurements, such as flow rates, phase fractions or velocities, see e.g., Yan et al. (2018), Alakeely and Horne (2021), Dang et al. (2019) and Li et al. (2021). Furthermore, in Lin et al. (2020), a deep learning model is used to predict different two-phase flow patterns in inclined pipes based on superficial velocities of the individual phases and inclination angles. Moreover, image processing techniques based on deep convolutional neural networks have been presented in Poletaev et al. (2020), Haas et al. (2020) and Cerqueira and Paladino (2021) for the detection, reconstruction, and analysis of gas bubbles in vertical pipes and micro-channels, for the recognition of flow patterns in micro pulsating heat pipes (Kamijima et al., 2020; Ahmad et al., 2022) as well as for the extraction of relevant water regions as pre-processing step for two-phase PIV-measurements in the field of ship and ocean engineering (Yu et al., 2021). For the quantification of separated and intermittent gas–liquid flow patterns in horizontal pipes, such as stratified wavy or slug flow, such advanced image processing techniques have not been reported. For this, image filter based methods are typically used, which are sensitive to changes in image quality, contrast, and recording set-up. The deep-learning based image processing technique presented here overcomes these problems to certain extent and provides a quantification for other separated flow pattern, namely, stratified, wavy, plug, and slug flow.

The proposed deep-learning model was trained in a supervised manner to correctly predict the liquid level time series from video recordings of gas–liquid flows. The model, a convolutional neural net, was extensively trained and tested with video data from real slug flow experiments and classifies each region in a video frame into its respective phase being either liquid or gas. For supervised learning the data has to be labelled. To do so, the respective video frames were used to create hand-labelled segmentation maps. Furthermore, the consistency of the hand-labelled data and the predictions of the trained model have been evaluated in an inter-observer reliability test. For further

evaluations of the reliability and versatility of the trained model, data from experiments are considered, which differ from the ones used for training and testing. These data also include an experiment with a different flow pattern, namely a stratified wavy flow.

2. Methods

In this section, the architecture of the convolution neural network as well as the experiments, the data, and its acquisition are described. Furthermore, it is explained, how the deep learning model is used with pre- and post-processing steps to extract the liquid level time series from the video data of the experiments. Moreover, the training and test procedure for the deep learning model is described. Finally, the accuracy and error metrics are provided that are used to evaluate the trained model.

2.1. Convolutional neural network

Deep convolutional neural networks are state-of-the-art machine learning techniques for image classification and segmentation problems (Dhillon and Verma, 2020; Aloysius and Geetha, 2017). In this paper, the liquid level extraction is considered as an image segmentation problem, where regions of liquid and gas need to be identified and segmented in the video frames. For this purpose, a specific convolutional neural network, the so-called U-net, was chosen. This architecture was introduced by Ronneberger et al. (2015) and has been successfully applied in many image-to-image learning problems, e.g., computer tomography, see Mao et al. (2016), Dosovitskiy et al. (2015) and Jin et al. (2017). The U-net is able to achieve accurate results with only few labelled training data and was therefore taken in this paper. The structure was altered from its original form and adapted from Sterbak (2018). The final architecture is illustrated in Fig. 2. An RGB-image input with the dimensions of 128×1024 px is passed through several layers of convolutions comprising a contracting part (left part of the U-shape), a bottleneck with minimal dimension in the centre of the U-shape and an expansive part (right part of the U-shape). The contracting part transfers the input image into a feature map with lower spatial dimensions but a higher number of feature channels. The expansive part generates the resulting segmentation map with the same spatial dimensions as the input image from the lower dimensional feature maps. In Fig. 2, the blue rectangles represent multi-channel feature maps. Here, their dimensions are also given, where the first two entries of the 3-tuple are related to the spatial dimensions (height and length) of the input image and the last entry corresponds to the number of feature channels. The basic operation in this network is the 3×3-convolution, which is followed by a batch normalization operation and a rectified linear unit (ReLU) as activation function. For the contracting part, the spatial dimensions are down-sampled with 2×2-max-pooling operations with stride 2. After each max pooling operation the number of features produced by the 3×3-convolution is doubled. For the expansive part, the feature-maps are up-sampled by 3×3-(up)-convolutions, which halve the number of feature channels, but double the number of spatial dimensions. For additional information in the reconstruction of the higher dimensional map, the feature maps from the corresponding level of the contracting part are copied and concatenated after each up-convolution. Finally, the segmentation map results from a 1×1-convolution and an activation operation using the Sigmoid function. The operations are illustrated as coloured arrows in Fig. 2. For details on the architecture and the operations, see Ronneberger et al. (2015) and Sterbak (2018).

2.2. Experiments

The used deep-learning model was trained and tested with data from video observations of horizontal gas–liquid flows. These flows are two-phase gas–oil or gas–water flows, as well as three-phase gas–oil–water flows, where the liquid phase is a homogeneous mixture of oil and water. The experiments were performed by TÜV SÜD NEL and DNV as part of the projects *Multiphase flow metrology in oil and gas production* (MultiFlowMet I) (Crawford, 2018) and *Multiphase flow reference metrology* (MultiFlowMet II) (Pieper, 2020).

The experimental set-ups are illustrated in Fig. 3. They consist of a horizontal inflow section followed by a vertical measurement section, but here the latter part is of minor interest since the flows in the horizontal pipe are investigated. For the set-up of the MultiFlowMet I project, the horizontal inflow section consists of a straight horizontal pipe with an inner diameter $D = 0.0972$ m and two different lengths $L_{\mathrm{inflow}} \in \{100D, 500D\}$, followed by a transparent Perspex viewing section with a length of $L_{\mathrm{viewSec}} = 5D$. For the set-up of the MultiFlowMet II project, the horizontal inflow section consists of a straight horizontal pipe with an inner diameter $D = 0.06664$ m and three different lengths $L_{\mathrm{inflow}} \in \{100D, 300D, 600D\}$, followed by a transparent Perspex viewing section with a length of $L_{\mathrm{viewSec}} = 9D$.

In Table 1, the laboratory that conducted the experiments, the recording set-up for the video data (see also Fig. 3), the fluids and their properties for some operating conditions are given for the two projects, for details see Crawford (2018) and Pieper (2020). In Table 2, the superficial velocities and lengths of the horizontal inflow sections L_{inflow} of the considered two- and three-phase flows of MultiFlowMet I and MultiFlowMet II are given, respectively. For all flows except the stratified wavy flow of Experiment Nr. 13, the slug flow pattern was observed. Please note that, for the stratified wavy flow experiment Nr. 13 and for the slug flow experiments Nr. 1–7 and Nr. 9–12, the interface is clearly visible from the side, which is necessary for the algorithm to work. These experiments have been considered in the training, testing, and evaluation procedure (see Sections 3.1 and 3.3). For the slug flow experiments Nr. 8, 14, 15, and 16, the gas–liquid interface is (partially) unrecognizable due to dispersed phenomena, such as foam or spray. These flows have been considered for the investigations on the limitations of the proposed image processing technique, see Section 4.

The flows were recorded at the viewing sections from the side using a high-speed RGB-camera with a frame rate of 240fps. For each experimental set-up, two different video recording set-ups were used, see Fig. 3. The set-ups 1, 2 and 3 were recorded at NEL, and set-up 4 was recorded at DNV. In set-up 1, the Paraflex oil has an orange–brown colour, the background is black and the viewing section is illuminated from the front. In set-up 2, the Brine water is of grey colour but slightly transparent, the background has a dark blue colour and the viewing section is illuminated from below. In set-up 3, the Paraflex oil (HT9) has a red-brown colour, the background is blue and the viewing section is illuminated from the front. And in set-up 4, the Exxsol oil (D120) is ocher-green, the background has a blue colour and the viewing section is illuminated from behind. Please note that nitrogen gas is

Table 1

Laboratory that conducted the experiments, the set-up for the video recordings (see Fig. 3) and fluid properties for some operating conditions that are used in the experiments.

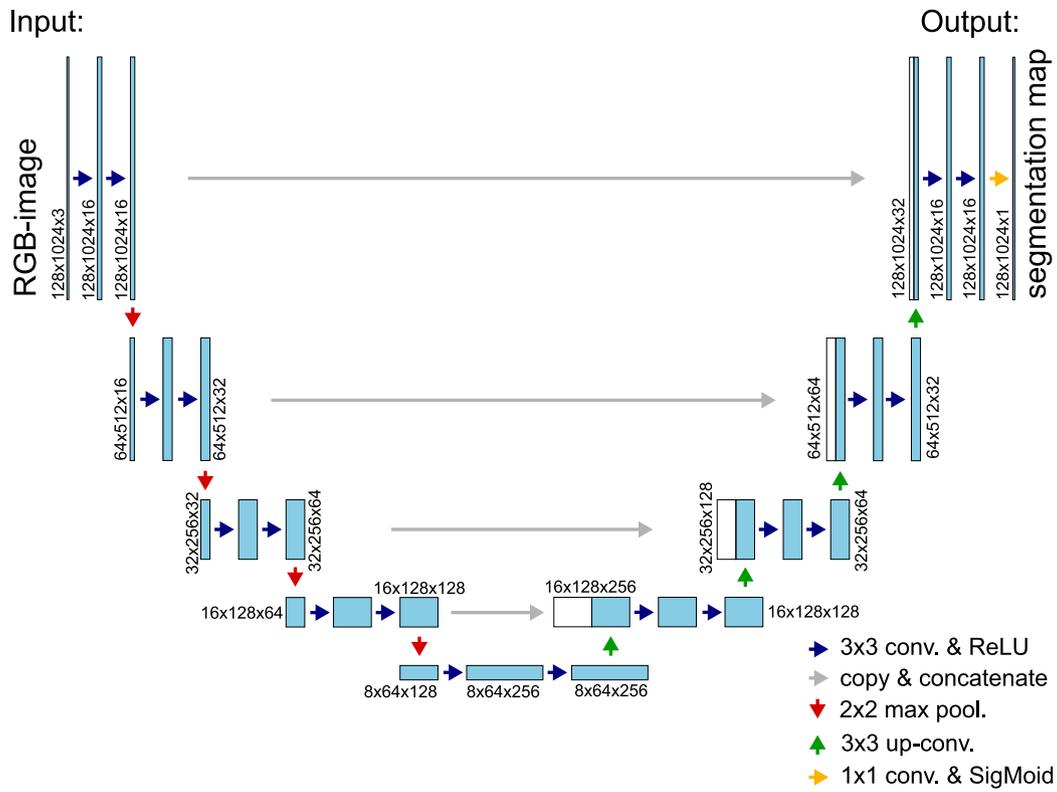| Lab	Set-up	Fluid	Density in kg m^{-3}	Viscosity in cP	Temp. in °C	Pressure in bar (g)
MultiFlowMet I						
NEL	1 & 2	Paraflex oil	831.40	16	20	9
		Brine Water	1037.89	1.15		
		Nitrogen gas	11.51	0.018		
MultiFlowMet II						
NEL	3	Paraflex HT9	815.10	7.32	45	9
		Brine Water	1024.19	0.65		
		Nitrogen gas	10.5	0.019		
DNV	4	Exxsol D120	827.55	5.10	18	8
		Brine Water	1031.69	1.04		
		Nitrogen gas	10.46	0.018		

Fig. 2. Illustration of the architecture of the U-net. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
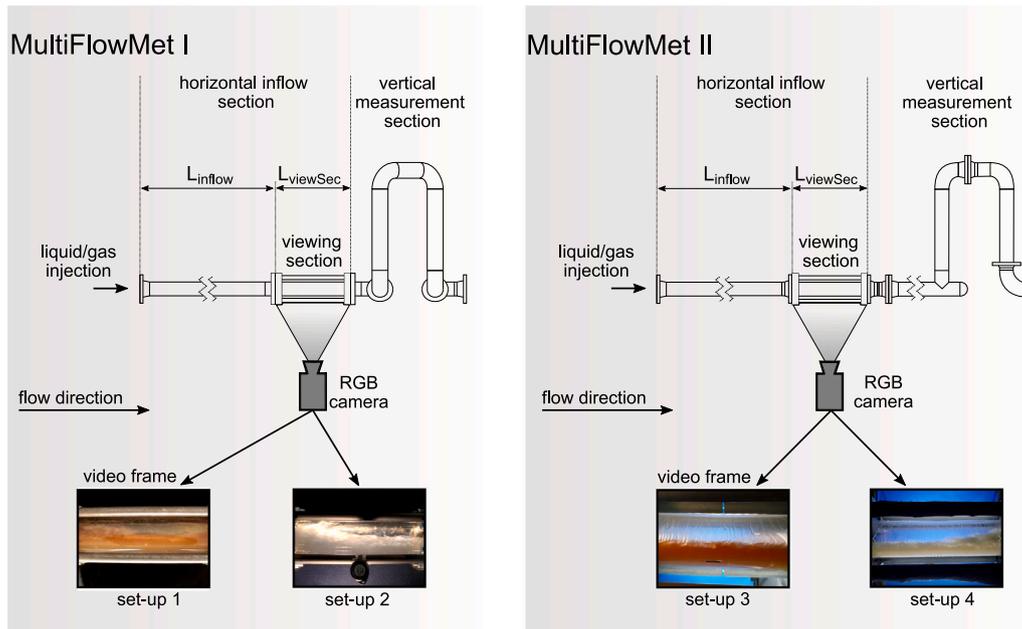

Fig. 3. Illustration of the experimental setting of the projects MultiFlowMet I and MultiFlowMet II with the different recording set-ups. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

transparent and colourless, similar to other common gases in the oil and gas industry, e.g., natural gas and argon (Pieper, 2020). Hence, the background colour is visible through the gas for all recording set-ups. Furthermore, for the three-phase gas–oil–water flows recorded in set-up 1, 3 and 4, the watercut is relatively small. Hence, the water does not form a separate liquid layer and the liquid phase appears in the video observations as a homogeneous oil–water mixture with similar colour as the oil. In Tables 3 and 6, the video recording set-up as well

as the recorded time (length of the videos) are given for the considered experiments of MutliFlowMet I and II.

2.3. Pre- and post-processing

The U-net, described in Section 2.1, is used to segment liquid and gas regions in parts of the video frames. However, before the U-net is applied, the video frames need to be prepared in several

Table 2
Superficial velocities $v_s^{\text{oil}}, v_s^{\text{water}}, v_s^{\text{gas}}$ and lengths of inflow section L_{inflow} in inner diameters D for the considered experiments of the MultiFlowMet I & II project.

Experiment	Superficial velocities in m s⁻¹			L_{inflow}
	v_s^{oil}	v_s^{water}	v_s^{gas}	in D
MultiFlowMet I				
Nr. 1	1.349	0	1.692	100, 500
Nr. 2	1.869	0	0.627	100
Nr. 3	1.930	0	0.751	100, 500
Nr. 4	2.547	0	3.185	100, 500
Nr. 5	0.986	0.328	2.849	100
Nr. 6	1.963	0.638	0.892	100
Nr. 7	0	1.847	0.678	100
Nr. 8	3.386	0	1.276	100
MultiFlowMet II				
Nr. 9	0.822	0	3.285	100, 300, 600
Nr. 10	1.218	0	1.179	100, 300, 600
Nr. 11	1.610	0	0.704	100, 300, 600
Nr. 12	2.390	0.006	2.398	100
Nr. 13	0.475	0	3.246	100
Nr. 14	2.352	1.559	1.548	100
Nr. 15	0.399	0	7.699	100
Nr. 16	0.739	0.470	11.442	100

pre-processing steps. Furthermore, to extract a time series from the gas–liquid segmentation maps of the U-net, additional post-processing steps are necessary.

In Fig. 4, the complete processing pipeline of the liquid level extraction from the video observations is illustrated. The first step is to extract a vertical line (pixelcolumn) through the pipe at a fixed x-position for every frame, i.e., time step, and stack it. From this procedure, an RGB-pixelcolumn over time is obtained, which represents the phase distribution along the observed vertical line through the pipe and its temporal changes at a fixed x-position. Because of this, the gas–liquid interface visible in this image is associated with the liquid level time series with respect to the inner pipe diameter. Under this construction, the frame rate of the video represents the sample rate of the time series. Therefore, the RGB-pixelcolumn over time provides the basis for further calculations. In the second step, the RGB-pixelcolumn over time is interpolated to a uniform height (y-component) of 128 px and cut into segments with a length (t-component) of 1024 px, to meet the input dimension criterion for the chosen U-net architecture. In the third step, the evenly sized segments are normalized to reduce the influence of disturbances in the video recording set-up, such as differences in luminance or colour. For this normalization, the *z-score* (also called statistical standardization or standard score, see Larsen and Marx (2012)) is applied RGB-component wise. This step completes the pre-processing and the standardized image segments are passed to the U-net to perform the segmentation. The output of the U-net is a continuous segmentation map with values in $[0, 1]$, where 1 (white) indicates gas and 0 (black) indicates liquid. In the first post-processing step, the continuous segmentation maps are binarized (with a threshold of 0.5) to obtain a sharp gas–liquid interface. Afterwards, the segments are concatenated in correct order to get a segmentation map for the complete RGB-pixelcolumn over time. In the last step, the vertical position of the gas–liquid interface (edge between black and white regions) is detected in the binarized and concatenated segmentation map. In case of multiple vertical interface positions at one time step, such as for bubbles or droplets, the values are averaged to get a unique representation of the interface over time. This ensures the property of a mathematically well-defined function for the extracted liquid level time series in case of multiple vertical interface positions such that it can be used for further time series analyses. It should be noted that this averaging can lead to a misrepresentation of certain flow structures in the liquid level time series. With this procedure the liquid level time series are obtained from the video recordings of multiphase flows. The

code for the extraction of the liquid level time series from the RGB-pixelcolumns over time is available as Jupyter Notebook in Olbrich et al. (2021c).

2.4. Training and testing

The U-net was trained and tested with data from horizontal gas–liquid slug flows. For this, the video data of 18 different slug flows were used. That includes data from 9 experiments of the MultiflowMet I project and 9 experiments of the MultiflowMet II project, specifically Nr. $1 - 6$, as well as Nr. $9 - 11$ for all inflow lengths, see Table 3. For the testing in the optimization process of the training procedure, a subset of the data is needed, which is disjoint to the training data. Here, the randomly chosen two experiments Nr. 5 - $100D$ and Nr. 11 - $300D$ are used for testing and the remaining 16 experiments are used for training. Please note that the naming of the individual experiments are given in the form of (Nr. - L_{inflow}).

To prepare the video data for the training of the U-net, the RGB-pixel-columns over time are extracted and normalized as described in Section 2.3 and Step $1 - 4$ in Fig. 4. Furthermore, binary segmentation masks are needed as reference in the training and testing process, which represent a correct classification into gas and liquid. These masks were generated from hand-labelled gas–liquid interfaces in the RGB-pixelcolumns over time for all experiments. They have a sharp interface with values of 1 for gas and 0 for liquid. Since they are extracted from the RGB-pixelcolumns over time, the masks represent an approximation of the temporally resolved gas volume fraction fields in a vertical line through the pipe at a fixed position. In the training and testing of the U-net, the masks are compared with the predicted segmentation maps to determine an accuracy for the prediction. For this, the masks are also transformed into evenly sized segments of 128×1024 px (see Step 2 in Fig. 4). This results in 483 pairs of evenly sized RGB-segments and corresponding mask segments for the training set as well as 61 of such pairs for the test set.

Since the segmentation includes only 2 classes, i.e., gas and liquid, the *binary accuracy* function was chosen as accuracy metric, see Eq. (2) and Chollet et al. (2015). Furthermore, for the training process the stochastic gradient descent optimization method *adaptive moment estimation (ADAM)* was set together with the *binary cross-entropy*-loss function and a dropout of 5%. Details can be found in Ronneberger et al. (2015), Sterbak (2018), Chollet et al. (2015) and Kingma and Ba (2017). For the training, the dropout-layer is located after every max-pooling operation in the contracting part of the U-net and after every concatenation-operation in the expansive part of the U-net, see Fig. 2. Furthermore, a mini batch size of 32 was used to train the U-net over a maximum number of 50 epochs. Early stopping (Prechelt, 1998) was applied. For details, see Sterbak (2018) and Chollet et al. (2015). The model was trained and tested using *python* version 3.7, *tensorflow* version 2.3 and *Keras* version 2.4.3, see Chollet et al. (2015) and Abadi et al. (2015). The implementation of the model, the code for the training of the model as Jupyter Notebook, and the weights of the trained model are available in Olbrich et al. (2021c).

2.5. Accuracy and error metrics

In this section, the metrics are given, which are used to evaluate the segmentation maps from the U-net as well as the extracted liquid level time series.

The binary accuracy of a predicted segmentation map A^{pred} and a corresponding mask A^{mask} is given by

$$acc_{\text{bin}}(A^{\text{pred}}, A^{\text{mask}}) = \frac{1}{N_t \cdot N_y} \sum_{i=1}^{N_t} \sum_{j=1}^{N_y} \phi\left(A_{ij}^{\text{pred}}, A_{ij}^{\text{mask}}\right), \qquad (2)$$
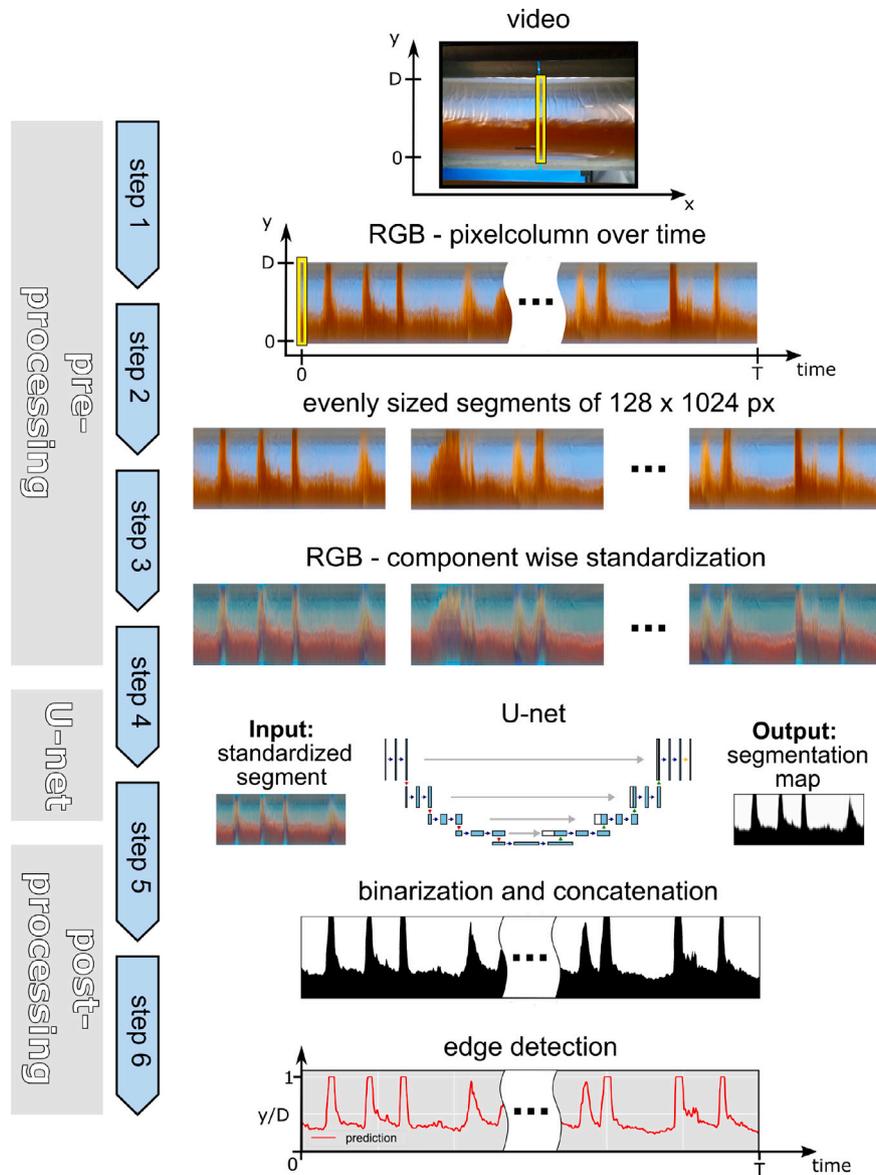
Fig. 4. Illustration of the processing pipeline from the raw video observations to the final liquid level time series. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where N_t and N_y denote the number of pixels of the mask in t-direction and y-direction, respectively, and the pixelwise binary evaluation function ϕ is given by

$$\phi(A_{ij}^{\text{pred}}, A_{ij}^{\text{mask}}) = \begin{cases} 1, & \text{if } A_{ij}^{\text{pred}} > 0.5 \wedge A_{ij}^{\text{mask}} > 0.5 \\ 1, & \text{if } A_{ij}^{\text{pred}} \leq 0.5 \wedge A_{ij}^{\text{mask}} \leq 0.5 \\ 0, & \text{else.} \end{cases} \quad (3)$$

Then, ϕ indicates if a pixel in the predicted segmentation map was successfully classified as gas or liquid. This function is pre-implemented in the open source software library *Keras*, see Chollet et al. (2015).

The deviation of the liquid level times series, which are extracted from the hand-labelled segmentation map h_L^{mask} and the U-net output h_L^{pred} can be measured in terms of the mean absolute error (MAE), given by

$$\epsilon(h_L^{\text{pred}}, h_L^{\text{mask}}) = \frac{1}{N_t} \sum_{i=1}^{N_t} \left| h_L^{\text{pred}}(t_i) - h_L^{\text{mask}}(t_i) \right|. \quad (4)$$

3. Results

In this section, the results of the liquid level extraction with the deep learning model are presented for the training and testing procedure, as well as for additional evaluations on different data sets. The results include the accuracy of the predicted segmentation maps with respect to the hand-labelled masks and the mean absolute error of the extracted liquid level time series. Furthermore, the consistency of the hand-labelled data and the prediction is investigated in an inter-observer reliability test. Moreover, the limitations of the proposed image processing technique are demonstrated and discussed.

3.1. Training and testing

The U-net was trained on a set of 483 pairs and tested on a set of 61 pairs of RGB-images from horizontal slug flow and corresponding masks, as described in Sections 2.3 and 2.4. The training was terminated after 42 epochs due to early stopping. The best model was found after 31 epochs. For this model, a mean binary accuracy of 97.91%
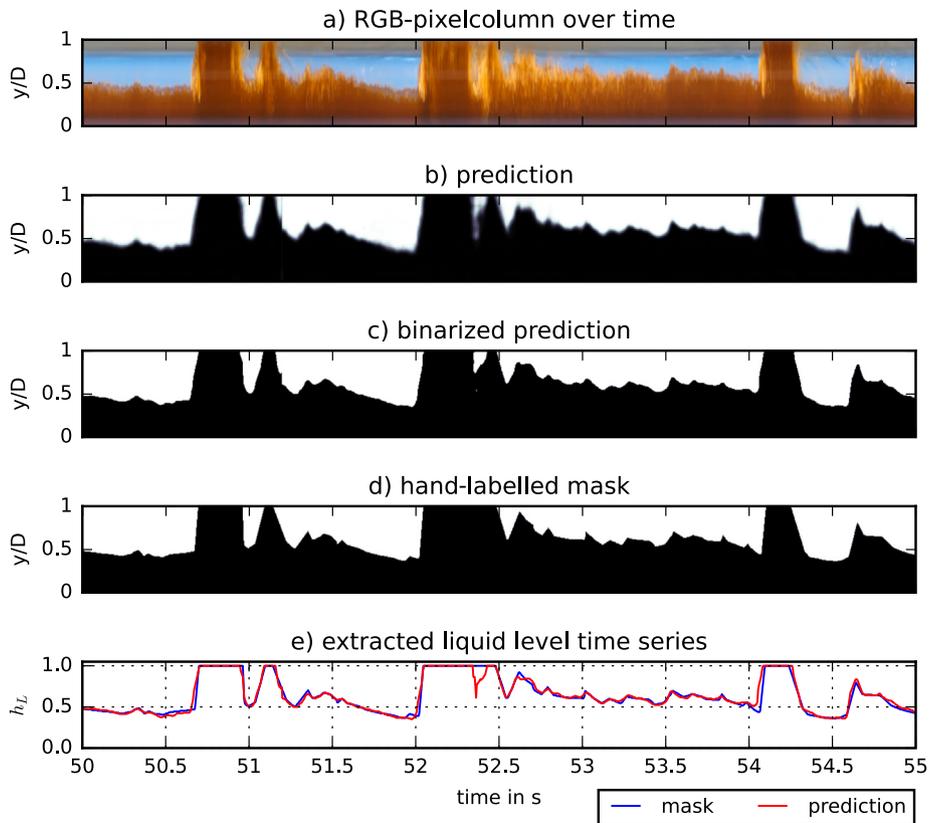
Fig. 5. Prediction of the U-net and extracted liquid level from Experiment Nr. 10 - $100D$ in comparison with hand-labelled mask for the interval $[50\,\text{s}, 55\,\text{s}]$.

for the training set and a mean binary accuracy of 97.74% for the test set, was achieved. These accuracies are high compared to the reported training and test accuracies in between 89% and 98.7% of other successfully trained deep-learning based gas and liquid segmentation models, see e.g. Cerqueira and Paladino (2021), Yu et al. (2021) and Ahmad et al. (2022).

In Fig. 5, the RGB-pixelcolumn over time from Experiment Nr. 10 - $100D$, the prediction of the U-net, its binarization, the corresponding hand-labelled mask, as well as the extracted liquid level time series are given. Please note that, Experiment Nr. 10 - $100D$ belongs to the training set. The prediction of the U-net and its binarization show good agreement with the mask. Furthermore, the gas and liquid regions are segmented in more detail in the prediction, compared to the hand-labelled mask, as can be seen for instance at the slug between $52\,\text{s}$ and $52.5\,\text{s}$ in Fig. 5. Here, the hand-labelled mask shows one larger slug, but the prediction shows two slugs, which are separated by a short gas bubble and foam. Due to the foamy areas between the rear of the first slug and the front of the second slug, this (optical) separation is not obvious. Nevertheless, the separation of the two slugs can be verified in the RGB-pixelcolumn over time, see Fig. 5a. Hence, in this case, the prediction from the trained model is more consistent and detailed compared to the hand-labelled mask on which it was trained.

In Table 3, the binary accuracy (see Eq. (2)) of the predictions and the masks in full length, as well as the mean absolute error (see Eq. (4)) of the extracted liquid level time series from prediction and mask are given for all experiments used for training and testing. The binary accuracy of the segmentation maps varies from 96.86% to 98.85% and the mean absolute error of the liquid level time series varies from 1.15% to 3.12%. From these high accuracies and low errors for the training and testing, it can be concluded, that the model performs well for the considered types of data. Furthermore, since the net accurately predicts the segmentation maps also for the two test sets that were not used in training, it is able to generalize to unseen data from both experimental set-ups, respectively.

Table 3

The number of the experiment, the recording set-up, the recorded time (length of the video), the belonging to training or test set, the binary accuracy $acc_{\text{bin}}(A^{\text{pred}}, A^{\text{mask}})$ (see Eq. (2)) of the predicted segmentation map and corresponding mask as well as the mean absolute error $\epsilon(h_L^{\text{pred}}, h_L^{\text{mask}})$ (see Eq. (4)) of the liquid level time series, extracted from prediction h_L^{pred} and mask h_L^{mask} are given for all experiments used for training and testing.

Experiment (Nr. - L_{inflow})	Set-up	Rec. time in s	Training or Test	acc_{bin} in %	ϵ in %
Nr. 1 - $100D$	1	122	Training	98.01	2.00
Nr. 1 - $500D$	1	150	Training	98.65	1.38
Nr. 2 - $100D$	1	150	Training	97.91	2.08
Nr. 3 - $100D$	1	150	Training	98.17	1.82
Nr. 3 - $500D$	1	150	Training	98.85	1.15
Nr. 4 - $100D$	1	150	Training	97.40	2.59
Nr. 4 - $500D$	1	150	Training	97.73	2.26
Nr. 6 - $100D$	1	150	Training	97.15	2.85
Nr. 9 - $100D$	3	122	Training	96.86	3.12
Nr. 9 - $300D$	3	126	Training	97.61	2.41
Nr. 9 - $600D$	3	121	Training	97.73	2.31
Nr. 10 - $100D$	3	122	Training	97.54	2.45
Nr. 10 - $300D$	3	125	Training	98.37	1.62
Nr. 10 - $600D$	3	123	Training	98.47	1.52
Nr. 11 - $100D$	3	122	Training	97.86	2.13
Nr. 11 - $600D$	3	123	Training	98.12	1.84
Nr. 11 - $300D$	3	123	Test	97.72	2.26
Nr. 5 - $100D$	1	150	Test	97.75	2.42

3.2. Inter-observer test

In this section, the inter-observer reliability is considered to evaluate the consistency of the hand-labelled data sets, which were used for the training of the U-net. For this, three independent observers have labelled the gas–liquid interface for the first $60\,\text{s}$ of four chosen experiments from the training and test set, namely Nr. 5 - $100D$, Nr. 6 - $100D$, Nr. 9 - $100D$ and Nr. 10 - $300D$ (see Table 3). Please note

Table 4
The Pearson-correlation values for the pairwise comparison of the different labels.

Experiment	$r(h_L^{obs_1}, h_L^{obs_2})$	$r(h_L^{obs_1}, h_L^{obs_3})$	$r(h_L^{obs_2}, h_L^{obs_3})$
Nr. 5 - $100D$	0.973	0.965	0.969
Nr. 6 - $100D$	0.920	0.981	0.880
Nr. 9 - $100D$	0.938	0.914	0.931
Nr. 10 - $300D$	0.985	0.981	0.977

Table 5
The Pearson-correlation values for the pairwise comparison of the prediction with the different labels.

Experiment	$r(h_L^{pred}, h_L^{obs_1})$	$r(h_L^{pred}, h_L^{obs_2})$	$r(h_L^{pred}, h_L^{obs_3})$
Nr. 5 - $100D$	0.970	0.970	0.962
Nr. 6 - $100D$	0.954	0.933	0.901
Nr. 9 - $100D$	0.970	0.962	0.939
Nr. 10 - $300D$	0.993	0.987	0.981

that, the U-net was trained and tested with labels from Observer 1. The inter-observer or inter-rater reliability (IRR) is defined as the degree of relationship between the labels of different observers that are operating independently (Kottner and Dassen, 2008; Tinsley and Weiss, 1975). Here, the IRR for the hand-labelled time series is quantified by the normalized cross-correlation coefficient. In the IRR-context, this coefficient is also referred to as *Pearson's r* or *Pearson-correlation*, see Kottner and Dassen (2008) and Berman (2016).

For every experiment, the labelled data sets of the different observers show a strong correlation to each other with values in between 0.88 and 0.985, see Table 4. This indicates a high degree of relation between the hand-labelled liquid level time series of the different observers. Hence, the hand-labelled liquid level time series show consistency and reliability among the different observers and are therefore suited for the training of the U-net. This correlation was calculated using the function *pearsonr* of python's SciPy module (Virtanen et al., 2020). Please note that, it returned *p*-values for the null-hypothesis significance testing of less then 0.001 for all cases. Furthermore, this analysis is applied to the prediction of the U-net to quantify the degree of relation between the prediction and the labels of the independent observers, see Xiao et al. (2017).

As given in Table 5, the Pearson-correlation values for the pairwise comparisons of the prediction and the labels of the different observers reach from 0.901 to 0.993 with *p*-values of less then 0.001 for all cases. This shows a strong correlation in a similar range as for the observers (see Table 4), and therefore, indicates a consistency between the predictions of the U-net and the labels of the independent observers.

In addition to the Pearson-correlation, the pointwise errors between the hand-labelled time series of the different observers as well as between the labels and the prediction of the U-net are considered. For a quantification of the error between the observers, the three time series of the pointwise errors between the different observers $|h_L^{obs_i}(t_k) - h_L^{obs_j}(t_k)|_{k=1,\dots,N_t}$ for $i \neq j \in \{1, 2, 3\}$ are ensemble-averaged (Walburn et al., 1983) to obtain one time series of the average pointwise error in between the observer for every experiment, i.e.,

$$\langle |h_L^{obs_i} - h_L^{obs_j}| \rangle (t_k) = \frac{1}{3} \sum_{i \neq j \in \{1,2,3\}} |h_L^{obs_i}(t_k) - h_L^{obs_j}(t_k)|, \text{ for every } t_k. \quad (5)$$

The same is done for the comparison of h_L^{pred} and $h_L^{obs_j}$ with $j \in \{1, 2, 3\}$. These ensemble-averages are shown as boxplots on the right side of Fig. 6. Here, the boxplots represent the distribution of these errors. It can be seen that, the statistical quantities (mean, median) and ranges (interquartile range and range between 5th and 95th percentile) of the errors are smaller for $\langle |h_L^{pred} - h_L^{obs_j}| \rangle$ compared to $\langle |h_L^{obs_i} - h_L^{obs_j}| \rangle$ for all experiments. On the left side of Fig. 6, the range in between the minima and maxima of the liquid level time series for every time point, labelled by the three observers, are given for the interval of 5 s. This range represents a tolerance in the observation of the gas–liquid

Table 6
The number of the experiment, the recording set-up, the recorded time (length of the video), the binary accuracy $acc_{bin}(A^{pred}, A^{mask})$ (see Eq. (2)) of the predicted segmentation map and corresponding mask as well as the mean absolute error $\epsilon(h_L^{pred}, h_L^{mask})$ (see Eq. (4)) of the liquid level time series, extracted from prediction h_L^{pred} and mask h_L^{mask} for the three additional independent evaluations.

Experiment (Nr. - L_{inflow})	Set-up	Rec. time in s	acc_{bin} in %	ϵ in %
Nr. 7 - $100D$	2	60	96.99	3.00
Nr. 12 - $100D$	4	60	95.41	4.92
Nr. 13 - $100D$	3	60	97.68	2.32

interface or the liquid level. In addition to this, the liquid level time series from the prediction of the U-net is superimposed. Here, it can be seen that, the predicted time series is near or in the tolerance range.

Altogether, the hand-labelled parameters of the different observers have a strong correlation as well as low pointwise errors between each other. Hence, this parameter shows consistency in between the observers and is therefore a reliable parameter for the training of the U-net. Furthermore, the comparison of the predicted liquid level time series with the observers show not only similarly strong correlation values, but also smaller statistical quantities of the considered pointwise errors. Hence, the predictions of the U-net also provide liquid level time series, which are consistent with respect to the different observers.

3.3. Further evaluations on different data sets

For further evaluations on the reliability and versatility of the trained model, data from three additional experiments are considered, which differ from the ones used for training and testing.

The model was trained and tested for Paraflex oil–nitrogen slug flows with black or blue background (recording set-up 1 and 3), see Fig. 3 and Section 2.4. In contrast to this, the flows considered in this section are either recorded in a different set-up with different fluids, or for a different flow pattern. They inlcude the brine water–nitrogen slug flow experiment Nr. 7 - $100D$ from recording set-up 2 with grey liquid colour and dark blue background, the Exxsol oil–nitrogen slug flow experiment Nr. 12 - $100D$ from recording set-up 4 with an ocher-green liquid colour, and the stratified wavy Paraflex oil–nitrogen flow experiment Nr. 13 - $100D$ from recording set-up 3. Please note that, also the lighting conditions and occurring reflections differ from the training and test set, e.g., reflections in the back of the pipe for Experiment Nr. 7 - $100D$ and white colour on top of the slugs for Experiment Nr. 12 - $100D$, see Fig. 7. These differences are causing changes in contrast and RGB-intensity values. Together with the different flow pattern, this leads to a change in conditions for the model, compared to the training and testing data.

In Table 6, the binary accuracy of the prediction and the corresponding hand-labelled masks as well as the mean absolute error of the extracted liquid level time series are given. For the Brine water–nitrogen slug flow of Experiment Nr. 7 - $100D$, the binary accuracy of 96.99% and the error value of 3% are in the same ranges as for the training and testing that are [96.86%, 98.85%] and [1.15%, 3.12%], respectively. The same holds for the stratified wavy flow of Experiment Nr. 13 - $100D$, with a binary accuracy of 97.68% and an error of 2.32%. The prediction for the slug flow experiment Nr. 12 - $100D$ did not achieve such high accuracy as the other experiments. Nevertheless, with a binary accuracy of 95.41% and an error of 4.92%, it is still close to the other values and a reasonable result.

In Fig. 7, the RGB-pixelcolumn over time with the extracted liquid levels from the prediction and the hand-labelled mask are given for a time interval of 5 s for experiment Nr. 7 - $100D$, Nr. 12 - $100D$ and Nr. 13 - $100D$. As it can be seen in Fig. 7(i), the liquid level from the prediction for the brine water - nitrogen slug flow experiment Nr. 7 - $100D$ shows a frequent underestimation in foamy areas in between shorter slugs.
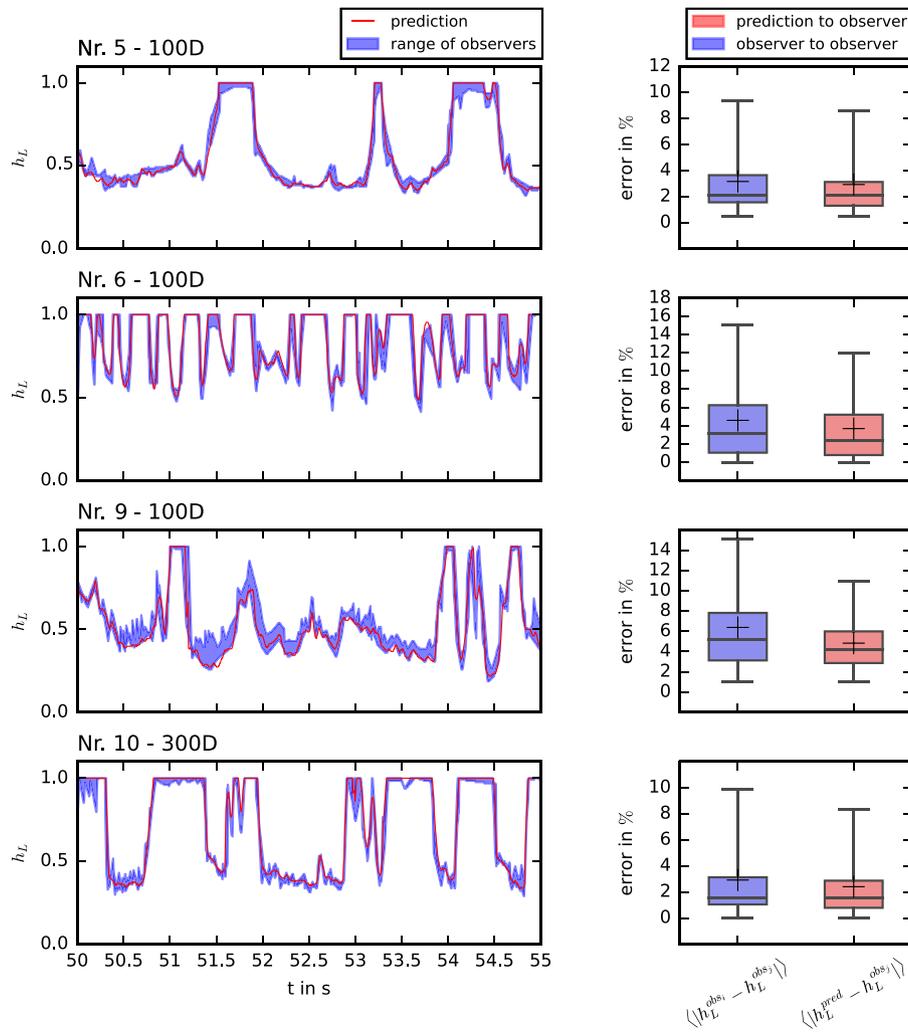
Fig. 6. Inter-observer comparison with prediction for the slug flow experiments Nr. 5 - 100D, Nr. 6 - 100D, Nr. 9 - 100D and Nr. 10 - 300D. **Left:** range of hand-labelled liquid level time series by the three independent observers and superimposed prediction for 5 s. **Right:** Boxplots of the ensemble-averaged point wise differences between the observers and between prediction and observers. Please note, the whiskers in the boxplots represents the 5th and 95th percentile, the coloured box represents the interquartile range (Q3-Q1), the horizontal line in the box represents the median, and the black cross represents the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

For the Exxsol oil - nitrogen slug flow experiment Nr. 12 - 100D in Fig. 7(ii), the prediction shows differences for slugs that are close to each other, see for instance the slugs at [50 s, 50.5 s] and [52 s, 52.5 s]. Furthermore, for this flow, the slug rears are predicted later compared to the mask. One reason for this is the liquid film that flows down on the inner walls of the pipe after a slug passed by. This also causes the top of the slugs to appear smeared out in the RGB-pixelcolumn over time and leads to the differences in between prediction and mask. Apart from the aforementioned deviations, the liquid level time series from the prediction and mask are in good agreement for all three flows.

In addition to this evaluation, the two previously unseen slug flow experiments Nr. 7 - 100D and Nr. 12 - 100D are further validated in the same manner as in the inter-observer test in Section 3.2. Therefore, the three independent observers have labelled the gas–liquid interface of these two slug flows. For the validation, the Pearson-correlation r between the different observers, as well as between the prediction and the observers, are considered, see Tables 7 and 8. Furthermore, the ensemble-averaged pointwise errors $\langle |h_L^{pred} - h_L^{obs_j}| \rangle$ and $\langle |h_L^{obs_i} - h_L^{obs_j}| \rangle$ are considered for this evaluation, see Fig. 8.

For the two additional experiments, the hand-labelled data sets of the different observers show a strong correlation to each other with values in between 0.922 and 0.974 and p-values of less then 0.001, see

Table 7
The Pearson-correlation values for the pairwise comparison of the different labels.

Experiment	$r(h_L^{obs_1}, h_L^{obs_2})$	$r(h_L^{obs_1}, h_L^{obs_3})$	$r(h_L^{obs_2}, h_L^{obs_3})$
Nr. 7 - 100D	0.974	0.958	0.953
Nr. 12 - 100D	0.941	0.942	0.922

Table 7. These values are similar to the ones obtained for the inter-observer test in Section 3.2 and indicate a high degree of relation between the hand-labelled liquid level time series of the different observers. The Pearson-correlation values for the pairwise comparisons of the prediction and the labels of the different observers reach from 0.868 to 0.967 with p-values of less then 0.001, see Table 8. This also shows a strong correlation and, therefore, indicates a consistency between the predictions of the U-net and the labels of the independent observers.

As mentioned in the discussion of Fig. 7, the predictions of the liquid level time series for Experiment Nr. 7 - 100D and Nr. 12 - 100D show some systematic deviations from the hand-labelled data. For the previously unseen nitrogen-water slug flow Experiment Nr. 7 - 100D, the prediction of the liquid level at the aerated liquid film region behind the slugs is often lower compared to the hand-labelled data. Moreover, for the previously unseen nitrogen–oil–water slug flow Experiment

a) RGB-pixelcolumn over time

b) extracted liquid level time series

(i) Experiment Nr. 7 - $100D$

a) RGB-pixelcolumn over time

b) extracted liquid level time series

(ii) Experiment Nr. 12 - $100D$

a) RGB-pixelcolumn over time

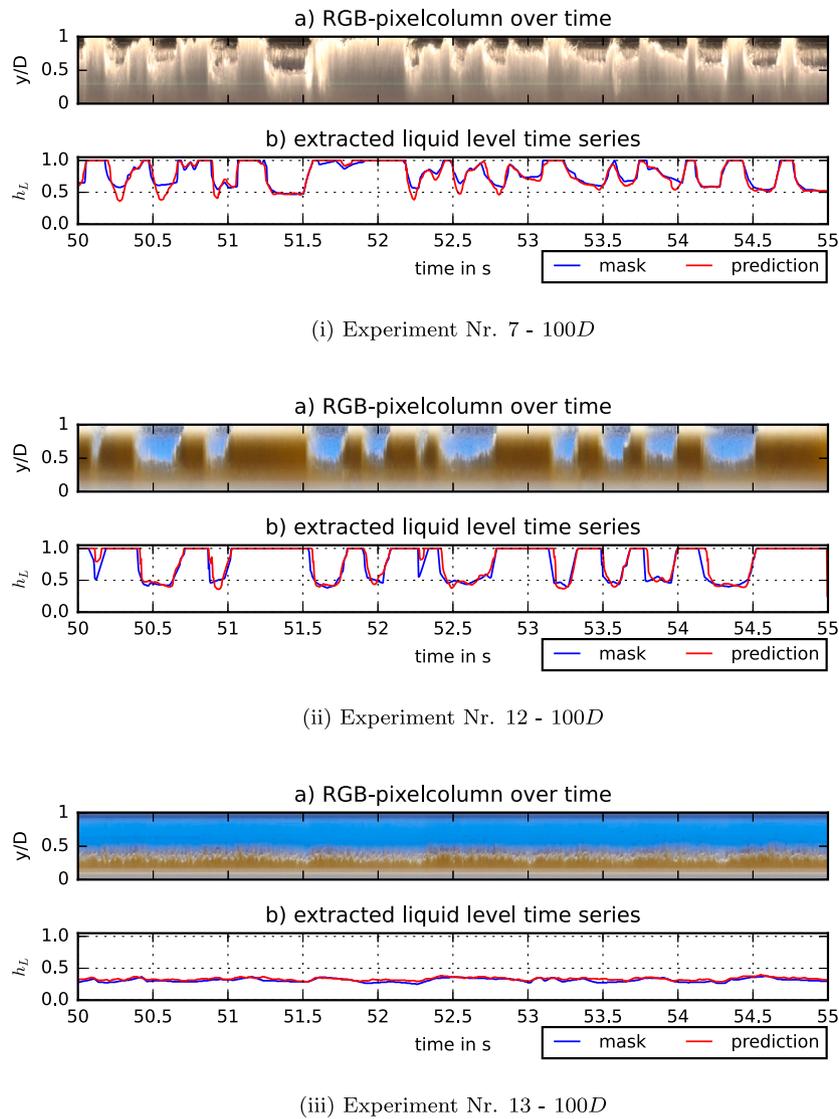b) extracted liquid level time series

(iii) Experiment Nr. 13 - $100D$

Fig. 7. Visualization of the results for three untrained flows and set-ups. For each experiment, the RGB-pixelcolumn over time and the extracted liquid level from the binarized prediction and the corresponding mask are given for the time interval $[50\,\text{s}, 55\,\text{s}]$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 8
The Pearson-correlation values for the pairwise comparison of the prediction with the different labels.

Experiment	$r(h_L^{\text{pred}}, h_L^{\text{obs}_1})$	$r(h_L^{\text{pred}}, h_L^{\text{obs}_2})$	$r(h_L^{\text{pred}}, h_L^{\text{obs}_3})$
Nr. 7 - $100D$	0.967	0.957	0.946
Nr. 12 - $100D$	0.905	0.900	0.868

Nr. 12 - $100D$, the slug rears often appear later in the prediction compared to the hand-labelled data. This is also shown in Fig. 8. This behaviour leads to slightly lower (but still high) Pearson-correlation values between the prediction and the observers for Experiment Nr. 12 - $100D$ compared to Experiment Nr. 7 - $100D$. Furthermore, this also leads to a slightly larger variation in the ensemble-averaged pointwise errors $\langle |h_L^{\text{pred}} - h_L^{\text{obs}_j}| \rangle$, compared to $\langle |h_L^{\text{obs}_i} - h_L^{\text{obs}_j}| \rangle$ for both experiments, see Fig. 8. Nevertheless, the liquid level predictions of both experiments are near or in the tolerance range of the observers, see left side of Fig. 8. Moreover, the statistical parameters of the ensemble-averaged pointwise errors between the predictions and the observers $\langle |h_L^{\text{pred}} - h_L^{\text{obs}_j}| \rangle$ are very similar to the ones between the different observers $\langle |h_L^{\text{obs}_i} - h_L^{\text{obs}_j}| \rangle$, see right side of Fig. 8. Hence, for the two

difficult unseen slug flow experiments Nr. 7 - $100D$ and Nr. 12 - $100D$, a strong consistency between the prediction and the observers can be concluded.

In the further, slug flow characteristics are calculated from the extracted liquid level time series for Experiment Nr. 12 - $100D$. The characteristics considered in this paper are the slug unit times T_u, the slug body times T_b, as well as their mean values, the mean slug frequency $\bar{f}_s$, the mean slug unit length $\bar{L}_u$, the mean slug body length $\bar{L}_b$, as well as the mean translational velocity of the slugs $\bar{u}_{\text{slug}}$, see Fig. 1 and Eq. (1). Considering these slug characteristics that can be obtained from the predicted time series exemplifies the physical insights provided by the results of the proposed image processing technique. It also allows further validation of the predicted time series. In that regard, it is shown that the predicted liquid level time series provide reasonable slug characteristics, also for the unseen data of Experiment Nr. 12 - $100D$ with the lowest accuracy and largest error values of all considered cases ($acc_{\text{bin}}(A^{\text{pred}}, A^{\text{mask}}) = 95.41\%$, $\epsilon(h_L^{\text{pred}}, h_L^{\text{mask}}) = 4.92\%$), see Tables 3 and 6.

In Fig. 9a, the predicted and hand-labelled liquid level time series of Experiment Nr. 12 - $100D$ are given for a time interval of $5\,\text{s}$, similar to Fig. 7(ii). For the calculation of the slug unit times T_u and slug body times T_b, a threshold value of 0.95 has been set to detect the slug fronts
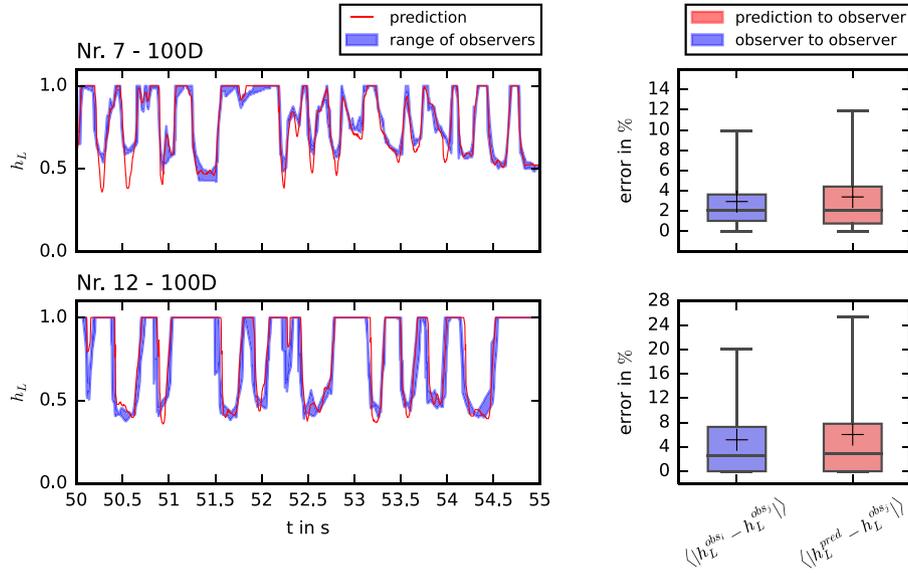
Fig. 8. Inter-observer comparison with prediction for the slug flow experiments Nr. 7 - 100D and Nr. 12 - 100D. **Left:** range of hand-labelled liquid level time series by the three independent observers and superimposed prediction for 5 s. **Right:** Boxplots of the ensemble-averaged point wise differences between the observers and between prediction and observers. Please note, the whiskers in the boxplots represents the 5th and 95th percentile, the coloured box represents the interquartile range (Q3-Q1), the horizontal line in the box represents the median, and the black cross represents the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
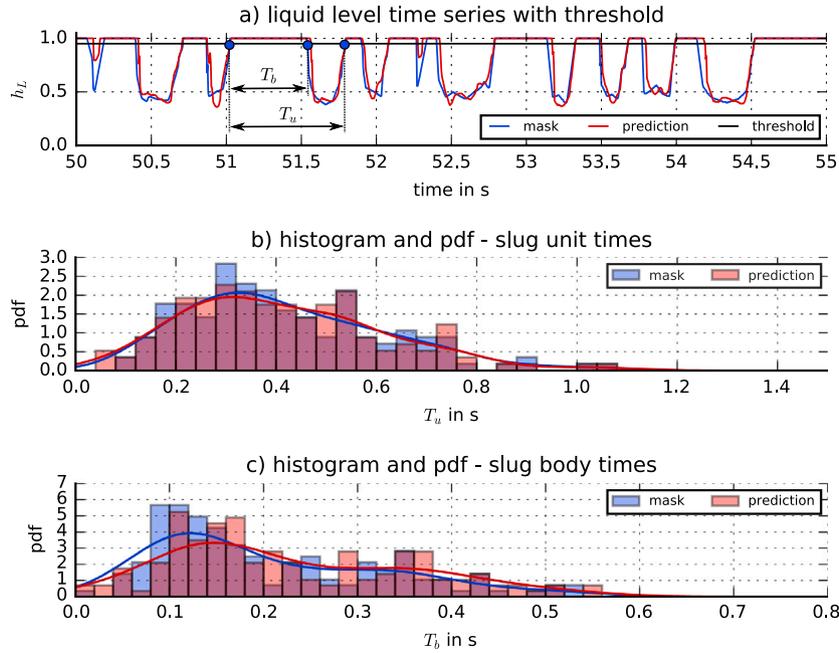

Fig. 9. The predicted and hand-labelled liquid level time series h_L^{mask} and h_L^{pred} for Experiment Nr. 12 - 100D with the threshold of 0.95 for slug detection (see a), histograms and probability density functions (pdf's) of the slug unit times T_u (see b) and slug body times T_b (see c) in comparison. Here, an illustration of the calculation of T_u and T_b is given in the time series plot (see a).

and slug rears in the time series, as illustrated in Fig. 9a. Please note that thresholding is the conventional procedure for this task, see Zhao et al. (2015), Baba et al. (2018) and Schmelter et al. (2021a), and therefore also applied in this investigation. Generally, the choice of the threshold values for the detection of slugs in the time series is not obvious and needs to be chosen individually for every time series. It should not be too high, otherwise larger slugs are separated by their entrained gas bubbles. However, it should also be chosen high enough to avoid the miscounting of large amplitude waves (Schmelter et al., 2021a). Furthermore, the histograms and probability density functions (pdf's) of the calculated slug unit times T_u and the slug body times T_b

are given in Fig. 9b and c for both, the hand-labelled and predicted liquid level time series. It can be seen that the predicted time series provide similar values for T_u and T_b compared to the hand-labelled time series. This is especially the case for the slug unit times T_u with nearly identical pdf's. The histograms and pdf's for the slug body times show a slight shift to larger values for the prediction compared to the hand-labelled time series. This is caused by the later prediction of the slug rears in the predicted liquid level time series, as already discussed for Fig. 7(ii). Furthermore, the mean slug frequency $\bar{f}_s$ can directly be obtained from the mean slug unit time $\bar{T}_u$, see Eq. (1),
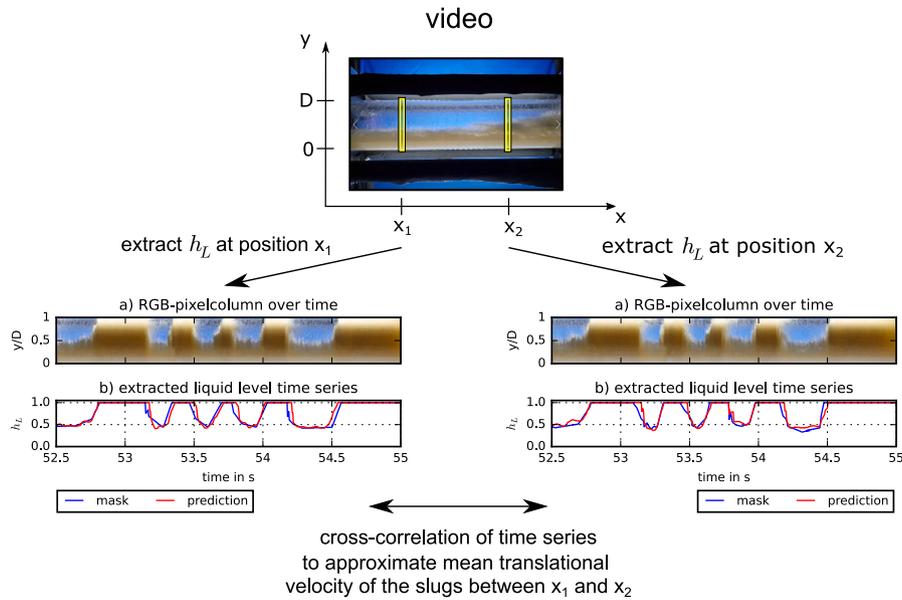
Fig. 10. Illustration of the calculation of the mean translational slug velocity $\bar{u}_{\text{slug}}$.

which is about $2.4\,\text{Hz}$ for both time series, see also Table 9. In addition, the mean translational velocity of the slugs is calculated for the hand-labelled and predicted liquid level time series. This was achieved by a cross-correlation analysis, which is typically used to calculate the mean translational velocity and approximate the length scales L_u and L_b of the slugs from hold-up and liquid level time series, see Baba et al. (2018), Viggiano et al. (2018) and Olbrich et al. (2021a). For this, the proposed image processing technique is applied to extract the liquid level time series at two different positions x_1 and x_2 along the pipe in the video with a distance of 2.2 inner diameters D, as illustrated in Fig. 10. For the comparison, also hand-labelled liquid level time series have been considered for these positions. Then, the time lags of the time series from position x_1 and x_2 are calculated by using the cross-correlation coefficient for the time series, resulting in 8 time steps $(0.033\,\text{s})$ for the prediction and 7 time steps $(0.029\,\text{s})$ for the hand-labelled time series with a sample rate of $240\,\text{Hz}$. Then, the mean translational velocity of the slugs $\bar{u}_{\text{slug}}$ is obtained by dividing the distance between x_1 and x_2 $(0.146\,\text{m})$ by the calculated time lag $(0.033\,\text{s}$ for the prediction and $0.029\,\text{s}$ for the hand-labelled time series), resulting in $\bar{u}_{\text{slug}} = 4.390\,\text{m s}^{-1}$ for the prediction and $\bar{u}_{\text{slug}} = 5.017\,\text{m s}^{-1}$ for the hand-labelled time series, see also Table 9.

Moreover, the mean slug body length L_b and the mean slug unit length L_u are approximated by multiplying the corresponding mean time scales $\bar{T}_u$ and $\bar{T}_b$ with the mean translational slug velocity $\bar{u}_{\text{slug}}$. The approximated length scales for both, the prediction and the hand-labelled time series can also be found in Table 9. Altogether, the slug flow characteristics obtained by the predicted liquid level time series are in good agreement with the ones from the hand-labelled time series. This holds in particular for the temporal scales $(\bar{T}_u, \bar{T}_b, \bar{f}_s)$. For the spatial scales $(\bar{L}_u$ and $\bar{L}_b)$, on the other hand, slight deviations between the predicted and the hand-labelled data can be observed due to the one time step difference in the lag detected in the cross-correlation procedure. Hence, this analysis also gives insight into how much error propagation plays a role if parameters are considered that are not directly determined but calculated from other derived quantities.

The characterization of slug flow with length and time scales as well as frequency spectra of the complete liquid level time series, obtained from the results of the proposed deep learning based image processing technique have already been used in the investigation of flow experiments and validation of numerical simulations, see Schmelter et al. (2021b).

Table 9
Comparison of the mean slug characteristics calculated from the hand-labelled and predicted liquid level time series.

Parameter	Mask	Prediction
Mean slug unit time $\bar{T}_u$ in s	0.421	0.415
Mean slug body time $\bar{T}_b$ in s	0.205	0.229
Mean slug frequency $\bar{f}_s$ in Hz	2.373	2.407
Mean transl. slug vel. $\bar{u}_{\text{slug}}$ in m s^{-1}	5.017	4.390
Mean slug unit length $\bar{L}_u$ in D	31.709	27.353
Mean slug body length $\bar{L}_b$ in D	15.392	15.056

Altogether, it can be concluded that the trained model can handle different types of data and provides reliable results for its specific task. Furthermore the proposed image processing technique provides accurate liquid level time series that allow a detailed characterization of the flow.

4. Limitations

The visual recognition of gas and liquid regions or the gas–liquid interface in the video data constitutes a major limitation of the successful extraction of the liquid level time series with the proposed deep learning based image processing technique. For flows, where the interface cannot be observed from the side, a meaningful gas–liquid segmentation cannot be provided by the trained model. This is for instance the case for dispersed or annular flow patterns as well as for flows with large amounts of liquid spray, e.g., for slug or wavy flows in the transition to a dispersed or annular flow pattern. Furthermore, for foamy/bubbly regions in the flow, the segmentation often includes foam into the liquid phase, leading to overestimated absolute values for the liquid level time series. This overestimation was investigated in detail in Olbrich et al. (2021b), where hand-labelled liquid level time series of slug flow have been compared to reference parameters of a conventional tomography measurement system.

To demonstrate the limitations of a meaningful liquid level extraction with the proposed image processing technique, the method is applied to four instances of slug flow with larger amounts of dispersed phenomena, such as foam/bubbles or spray/mist. These are Experiment Nr. 8 and 14 with high liquid and low gas flow rates leading to a high liquid level in the pipe with large amounts of short slugs and foam/bubbles (see Tables. 2, 10 and Figs. 11i, 11iii) as well as Experiment
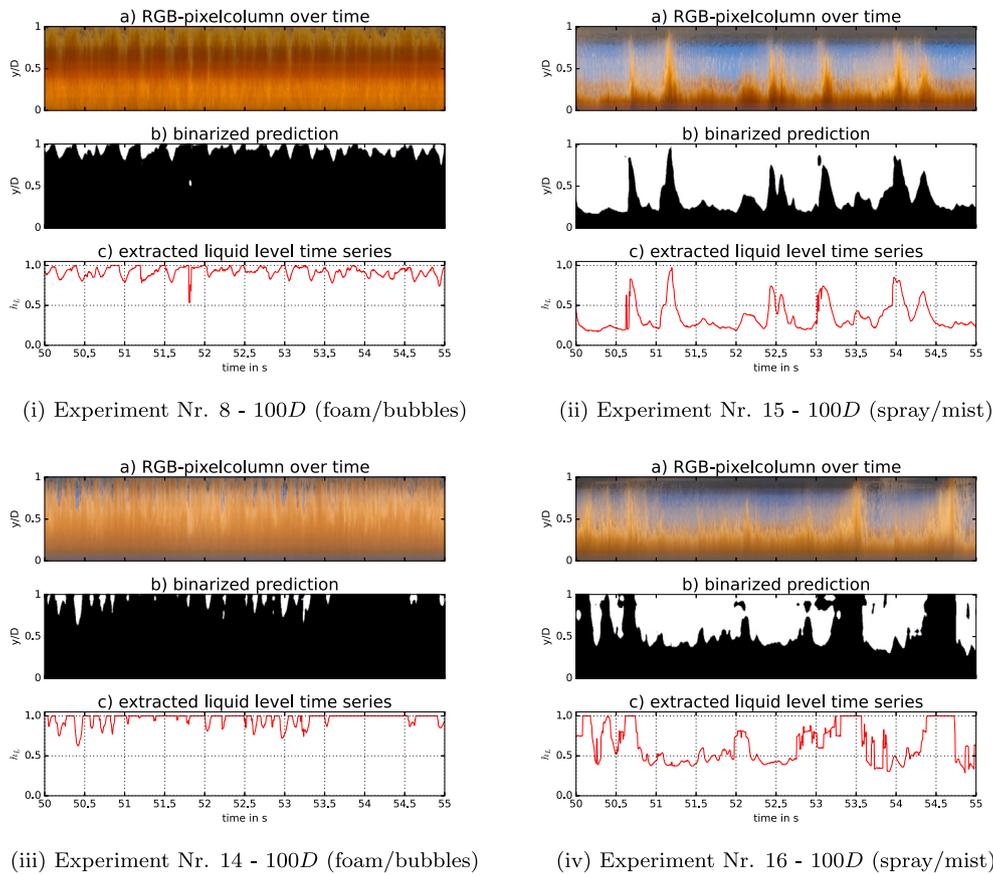
(i) Experiment Nr. 8 - $100D$ (foam/bubbles)

(ii) Experiment Nr. 15 - $100D$ (spray/mist)

(iii) Experiment Nr. 14 - $100D$ (foam/bubbles)

(iv) Experiment Nr. 16 - $100D$ (spray/mist)

Fig. 11. Visualization of the limitations of the prediction for four slug flows with large amounts of dispersed phenomena. For each experiment, the RGB-pixelcolumn over time (see a), the binarized predicted segmentation map (see b) and the extracted liquid level time series (see c) are given for the time interval $[50\,\text{s}, 55\,\text{s}]$.

Table 10
The number of the experiment, the recording set-up, the recorded time (length of the video), the superficial liquid and gas velocities $v_s^{\text{liquid}}, v_s^{\text{gas}}$ (see also Table 2) and the dominant dispersed phenomena in the flow leading to an unrecognizable gas–liquid interface.

Experiment (Nr. - L_{inflow})	Set-up	Rec. time in s	v_s^{liquid} in m s^{-1}	v_s^{gas} in m s^{-1}	Dispersed phenomena
Nr. 8 - $100D$	1	60	3.386	1.276	Foam/bubbles
Nr. 14 - $100D$	3	60	3.911	1.548	Foam/bubbles
Nr. 15 - $100D$	3	60	0.399	7.699	Spray/mist
Nr. 16 - $100D$	3	60	1.209	11.442	Spray/mist

Nr. 15 and 16 with high gas and low liquid flow rates leading to a low liquid level with fewer shorter slugs and large amounts of spray/mist (see Tables. 2, 10 and Figs. 11ii, 11iv).

Due to the dispersed phenomena, the gas–liquid interface becomes (at least partially) unrecognizable for the observer in the considered image data. Thus, a meaningful hand-labelled segmentation map as ground truth for a validation of the prediction is not obtainable. However, the predictions are presented with the RGB-pixelcolumns over time in order to give an impression about the limitations of the proposed deep-learning based image processing technique.

For Experiment Nr. 8 and 15, the binarized predictions show artefacts caused by the foam/bubbles or the spray/mist, see for instance the segmented bubble in the foam/bubbles at $[51.5\,\text{s}, 52\,\text{s}]$ in Fig. 11i(b) as well as the segmented lump of liquid in the spray/mist at $53\,\text{s}$ in Fig. 11ii(b). These artefacts lead to deviations for the extraction of the liquid level time series. Despite the fact that a detailed validation with hand-labelled segmentation masks cannot be made, the extracted liquid level time series show the dominant liquid structures in the flow and can therefore be used for a quantification of the slugs in

these two cases. For Experiment Nr. 14 and 16, on the other hand, with more dispersed phenomena compared to Experiment Nr. 8 and 15, the binarized predictions show larger areas of segmented liquid compared to what can visually be observed in the corresponding RGB-pixelcolumns over time (see 11iii(a,b) and 11iv(a,b)). This includes foam/bubbles or spray/mist, which are identified as liquid in the gas–liquid segmentation by the trained deep learning model, leading to impractical liquid level time series approximations.

5. Conclusions

In this paper, an image processing method based on a supervised deep learning model was presented, which extracts the liquid level time series from video recordings of liquid–gas flows in horizontal pipes. This method consists of a certain type of a deep convolutional neural network, called U-net, and several pre- and post-processing steps. The U-net was trained and tested with video data from horizontal oil–gas slug flows for the task of segmenting liquid and gas regions in the video frame data. For further evaluations of this model, additional independent video data were considered, which show different fluids, recording set-ups, and flow pattern. It was shown that, the trained model provides an accurate segmentation of oil and gas in the video data, even for previously unseen video recordings. In that regard, the model has proven to be versatile and is also applicable for other transparent gases. Furthermore, the extracted liquid level time series from the predicted segmentation maps show low errors. For the quantification of accuracy and error values, hand-labelled data was used as reference. The consistency between these hand-labelled data and the predictions of the U-net was shown in an inter-observer reliability test. Moreover, it was demonstrated how flow characteristics can be obtained from the results of the deep-learning based image processing technique.

Altogether, the presented method accurately extracts the liquid level time series from the considered video data. It can handle different types of data, even unseen data sets. Furthermore, it can overcome various noise effects, which are generally included in such image or video data. Once, the net is successfully trained, it predicts highly accurate segmentation maps in very short time. Prospectively, this method can provide parameters for the analysis and characterization of certain types of multiphase flows, in particular for the wavy or slug flow pattern, where temporal and spatial scales of the slugs, waves and bubbles, as well as their translational velocities can be derived from the extracted liquid level time series. The achieved characterization helps to assess and quantify problems for industrial operations that are induced by specific flow patterns. In addition, the proposed model has the potential for a segmentation of more complex three-phase flows, such as gas–oil–water flows with separate phases, on condition that it is trained with such data. Moreover, the proposed deep-learning based image processing technique can be used for monitoring multiphase flows for operation control if a transparent viewing section can be installed. This includes academic investigations under laboratory conditions as well as industrial applications, e.g., transportation pipelines in the oil and gas industry as well as cooling systems in the nuclear energy sector. In that regard, the link between the extracted liquid level time series and controlling parameters, such as pressure, flow rates, and phase ratio can be subject to further research.

CRediT authorship contribution statement

Marc Olbrich: Conceptualization, Methodology, Software, Investigation, Formal analysis, Validation, Visualization, Writing – original draft. **Leili Riazy:** Conceptualization, Methodology, Software, Investigation, Writing – review & editing. **Tobias Kretz:** Methodology, Software, Writing – review & editing. **Terri Leonard:** Investigation, Writing – review & editing. **Dennis S. van Putten:** Investigation, Writing – review & editing. **Markus Bär:** Supervision, Writing – review & editing. **Kilian Oberleithner:** Supervision, Writing – review & editing. **Sonja Schmelter:** Validation, Investigation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The Dataset of Experiment Nr. 11 and the source code related to this article can be found at https://gitlab1.ptb.de/mfm2/liquid_level_extraction_unet, an open-source online GitLab repository hosted at Physikalisch-Technische Bundesanstalt (PTB) (Olbrich et al., 2021c).

Acknowledgements

Dataset

The Dataset of Experiment Nr. 11 and the source code related to this article can be found at https://gitlab1.ptb.de/mfm2/liquid_level_extraction_unet, an open-source online GitLab repository hosted at Physikalisch-Technische Bundesanstalt (PTB) (Olbrich et al., 2021c).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. https://www.tensorflow.org/. Accessed: 2021-05-31, Software available from tensorflow.org.

Ahmad, H., Kim, S.K., Park, J.H., Jung, S.Y., 2022. Development of two-phase flow regime map for thermally stimulated flows using deep learning and image segmentation technique. Int. J. Multiph. Flow. 146, 103869. http://dx.doi.org/10.1016/j.ijmultiphaseflow.2021.103869, URL: https://www.sciencedirect.com/science/article/pii/S0301932221002895.

Al-Kayiem, H.H., Mohmmed, A.O., Al-Hashimy, Z.I., Time, R.W., 2017. Statistical assessment of experimental observation on the slug body length and slug translational velocity in a horizontal pipe. Int. J. Heat Mass Transfer 105, 252–260.

Al-Safran, E., 2009. Investigation and prediction of slug frequency in gas/liquid horizontal pipe flow. J. Pet. Sci. Eng. 69 (1), 143–155.

Alakeely, A., Horne, R., 2021. Application of deep learning methods to estimate multiphase flow rate in producing wells using surface measurements. J. Pet. Sci. Eng. 205, 108936.

Aloysius, N., Geetha, M., 2017. A review on deep convolutional neural networks. In: 2017 International Conference on Communication and Signal Processing. ICCSP, pp. 0588–0592.

Baba, Y.D., Aliyu, A.M., Archibong, A.E., Abdulkadir, M., Lao, L., Yeung, H., 2018. Slug length for high viscosity oil-gas flow in horizontal pipes: Experiments and prediction. J. Pet. Sci. Eng. 165, 397–411.

Baghernejad, Y., Hajidavalloo, E., Hashem Zadeh, S.M., Behbahani-Nejad, M., 2019. Effect of pipe rotation on flow pattern and pressure drop of horizontal two-phase flow. Int. J. Multiph. Flow. 111, 101–111.

Bao, H., Feng, J., Dinh, N., Zhang, H., 2020. Computationally efficient CFD prediction of bubbly flow using physics-guided deep learning. Int. J. Multiph. Flow. 131, 103378.

Berman, J.J., 2016. Chapter 4 - Understanding your data. In: Berman, J.J. (Ed.), Data Simplification. Morgan Kaufmann, Boston, pp. 135–187.

Cerqueira, R.F., Paladino, E.E., 2021. Development of a deep learning-based image processing technique for bubble pattern recognition and shape reconstruction in dense bubbly flows. Chem. Eng. Sci. 230, 116163. http://dx.doi.org/10.1016/j.ces.2020.116163, URL: https://www.sciencedirect.com/science/article/pii/S0009250920306953.

Chollet, F., et al., 2015. Keras. https://keras.io. Accessed: 2021-05-31.

Crawford, D., 2018. Final Publishable JRP Report and Associated Annex A - Multiphase Flow Metrology in Oil and Gas Production (ENG58). Technical Report, EURAMET.

Dang, W., Gao, Z., Hou, L., Lv, D., Qiu, S., Chen, G., 2019. A novel deep learning framework for industrial multiphase flow characterization. IEEE Trans. Ind. Inf. 15 (11), 5954–5962.

Dhillon, A., Verma, G.K., 2020. Convolutional neural network: a review of models, methodologies and applications to object detection. Prog. Artif. Intell. 9, 85–112.

do Amaral, C., Alves, R., da Silva, M., Arruda, L., Dorini, L., Morales, R., Pipa, D., 2013. Image processing techniques for high-speed videometry in horizontal two-phase slug flows. Flow Meas. Instrum. 33, 257–264.

Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., van der Smagt, P., Cremers, D., Brox, T., 2015. FlowNet: Learning optical flow with convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision. ICCV.

Drury, R., Hunt, A., Brusey, J., 2019. Identification of horizontal slug flow structures for application in selective cross-correlation metering. Flow Meas. Instrum. 66, 141–149.

Dukler, A.E., Fabre, J., 1994. Gas-liquid slug flow. Multiphase Sci. Technol. 8 (1–4), 355–469.

Elliott, A., Falcone, G., van Putten, D., Leonard, T., Haukalid, K., Pinguet, B., 2021. Investigating reproducibility in multiphase flow metrology: Results from an intercomparison of laboratories. Flow Meas. Instrum. 79, 101951.

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., Dehmer, M., 2020. An introductory review of deep learning for prediction models with big data. Front. Artif. Intell. 3, 4.

Graves, A., Mohamed, A.-r., Hinton, G., 2013. Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 6645–6649.

Haas, T., Schubert, C., Eickhoff, M., Pfeifer, H., 2020. BubCNN: Bubble detection using Faster RCNN and shape regression network. Chem. Eng. Sci. 216, 115467. http://dx.doi.org/10.1016/j.ces.2019.115467, URL: https://www.sciencedirect.com/science/article/pii/S0009250919309571.

Hanratty, T.J., 2013. Physics of Gas-Liquid Flows. Cambridge University Press.

Jin, K.H., McCann, M.T., Froustey, E., Unser, M., 2017. Deep convolutional neural network for inverse problems in imaging. IEEE Trans. Image Process. 26 (9), 4509–4522.

Kamijima, C., Yoshimoto, Y., Abe, Y., Takagi, S., Kinefuchi, I., 2020. Relating the thermal properties of a micro pulsating heat pipe to the internal flow characteristics via experiments, image recognition of flow patterns and heat transfer simulations. Int. J. Heat Mass Transfer 163, 120415. http://dx.doi.org/10.1016/j.ijheatmasstransfer.2020.120415, URL: https://www.sciencedirect.com/science/article/pii/S0017931020333512.

Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. arXiv:1412.6980.

Kottner, J., Dassen, T., 2008. Interpreting interrater reliability coefficients of the Braden scale: A discussion paper. Int. J. Nurs. Stud. 45 (8), 1238–1246.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60 (6), 84–90.

Larsen, R.J., Marx, M.L., 2012. Introduction to Mathematical Statistics and Its Applications, 5th Edition, Vol. 5. Prentice Hall.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Li, J., Hu, D., Chen, W., Li, Y., Zhang, M., Peng, L., 2021. CNN-based volume flow rate prediction of oil–gas–water three-phase intermittent flow from multiple sensors. Sensors 21 (4), 1245.

Lin, Z., Liu, X., Lao, L., Liu, H., 2020. Prediction of two-phase flow patterns in upward inclined pipes via deep learning. Energy 210, 118541.

Ma, M., Lu, J., Tryggvason, G., 2015. Using statistical learning to close two-fluid multiphase flow equations for a simple bubbly system. Phys. Fluids 27 (9), 092101.

Mao, X.-J., Shen, C., Yang, Y.-B., 2016. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS '16, Curran Associates Inc., Red Hook, NY, USA, pp. 2810–2818.

Mohmmed, A.O., Al-Kayiem, H.H., Nasif, M.S., Time, R.W., 2019. Effect of slug flow frequency on the mechanical stress behavior of pipelines. Int. J. Press. Vessels Pip. 172, 1–9.

Olbrich, M., Bär, M., Oberleithner, K., Schmelter, S., 2021a. Statistical characterization of horizontal slug flow using snapshot proper orthogonal decomposition. Int. J. Multiph. Flow. 134, 103453.

Olbrich, M., Hunt, A., Leonard, T., van Putten, D.S., Bär, M., Oberleithner, K., Schmelter, S., 2021b. Comparing temporal characteristics of slug flow from tomography measurements and video observations. Meas.: Sensors 18, 100222.

Olbrich, M., Riazy, L., Kretz, T., Schmelter, S., 2021c. Liquid-gas interface selection algorithm (LISA) with unet. https://gitlab1.ptb.de/mfm2/liquid_level_extraction_unet.

Olbrich, M., Schmeyer, E., Bär, M., Sieber, M., Oberleithner, K., Schmelter, S., 2020. Identification of coherent structures in horizontal slug flow. Flow Meas. Instrum. 76, 101814.

Olbrich, M., Schmeyer, E., Riazy, L., Oberleithner, K., Bär, M., Schmelter, S., 2018. Validation of simulations in multiphase flow metrology by comparison with experimental video observations. J. Phys. Conf. Ser. 1065, 092015.

Pedersen, S., Durdevic, P., Yang, Z., 2017. Challenges in slug modeling and control for offshore oil and gas productions: A review study. Int. J. Multiph. Flow. 88, 270–284.

Pieper, A., 2020. Final Publishable Report - Multiphase Flow Reference Metrology (16ENG07). Technical Report, EURAMET.

Poletaev, I., Tokarev, M.P., Pervunin, K.S., 2020. Bubble patterns recognition using neural networks: Application to the analysis of a two-phase bubbly jet. Int. J. Multiph. Flow. 126, 103194.

Prechelt, L., 1998. Early stopping - but when? In: Orr, G.B., Müller, K.-R. (Eds.), Neural Networks: Tricks of the Trade. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 55–69.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp. 234–241.

Schmelter, S., Knotek, S., Olbrich, M., Fiebach, A., Bär, M., 2021a. On the influence of inlet perturbations on slug dynamics in horizontal multiphase flow—a computational study. Metrologia 58 (1), 014003.

Schmelter, S., Olbrich, M., Knotek, S., Bär, M., 2021b. Analysis of multiphase flow simulations and comparison with high-speed video observations. Meas.: Sensors 18, 100154. http://dx.doi.org/10.1016/j.measen.2021.100154, URL: https://www.sciencedirect.com/science/article/pii/S2665917421001173.

Sterbak, T., 2018. U-Net for segmenting seismic images with keras. https://www.depends-on-the-definition.com/unet-keras-segmenting-images. Accessed: 2021-05-25.

Taitel, Y., Dukler, A., 1977. A model for slug frequency during gas-liquid flow in horizontal and near horizontal pipes. Int. J. Multiph. Flow. 3 (6), 585–596.

Tinsley, H.E., Weiss, D.J., 1975. Interrater reliability and agreement of subjective judgments. J. Couns. Psychol. 22 (4), 358–376.

Viggiano, B., Skjæraasen, O., Schümann, H., Tutkun, M., Cal, R.B., 2018. Characterization of flow dynamics and reduced-order description of experimental two-phase pipe flow. Int. J. Multiph. Flow. 105, 91–101. http://dx.doi.org/10.1016/j.ijmultiphaseflow.2018.03.020, URL: https://www.sciencedirect.com/science/article/pii/S0301932217309941.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods 17, 261–272.

Walburn, F.J., Sabbah, H.N., Stein, P.D., 1983. An experimental evaluation of the use of an ensemble average for the calculation of turbulence in pulsatile flow. Ann. Biomed. Eng. 11 (5), 385–399.

Wang, Y., Lin, G., 2020. Efficient deep learning techniques for multiphase flow simulation in heterogeneous porousc media. J. Comput. Phys. 401, 108968.

Xiao, S., Bucher, F., Wu, Y., Rokem, A., Lee, C.S., Marra, K.V., Fallon, R., Diaz-Aguilar, S., Aguilar, E., Friedlander, M., Lee, A.Y., 2017. Fully automated, deep learning segmentation of oxygen-induced retinopathy images. JCI Insight 2 (24).

Yan, Y., Wang, L., Wang, T., Wang, X., Hu, Y., Duan, Q., 2018. Application of soft computing techniques to multiphase flow measurement: A review. Flow Meas. Instrum. 60, 30–43.

Yu, C., Luo, H., Bi, X., Fan, Y., He, M., 2021. An effective convolutional neural network for liquid phase extraction in two-phase flow PIV experiment of an object entering water. Ocean Eng. 237, 109502. http://dx.doi.org/10.1016/j.oceaneng.2021.109502, URL: https://www.sciencedirect.com/science/article/pii/S0029801821009021.

Zhao, Y., Lao, L., Yeung, H., 2015. Investigation and prediction of slug flow characteristics in highly viscous liquid and gas flows in horizontal pipes. Chem. Eng. Res. Des. 102, 124–137. http://dx.doi.org/10.1016/j.cherd.2015.06.002, URL: https://www.sciencedirect.com/science/article/pii/S0263876215002075.