



# Practical object and flow structure segmentation using artificial intelligence

Ali R. Khojasteh<sup>1</sup> · Willem van de Water<sup>1</sup> · Jerry Westerweel<sup>1</sup>

Received: 30 April 2024 / Revised: 28 June 2024 / Accepted: 2 July 2024 / Published online: 30 July 2024  
© The Author(s) 2024

## Abstract

This paper explores integrating artificial intelligence (AI) segmentation models, particularly the Segment Anything Model (SAM), into fluid mechanics experiments. SAM's architecture, comprising an image encoder, prompt encoder, and mask decoder, is investigated for its application in detecting and segmenting objects and flow structures. Additionally, we explore the integration of natural language prompts, such as BERT, to enhance SAM's performance in segmenting specific objects. Through case studies, we found that SAM is robust in object detection in fluid experiments. However, segmentations related to flow properties, such as scalar turbulence and bubbly flows, require fine-tuning. To facilitate the application, we have established a repository ([https://github.com/AliRKhojasteh/Flow\\_segmentation](https://github.com/AliRKhojasteh/Flow_segmentation)) where models and usage examples can be accessed.

## 1 Introduction

Experiments in fluid mechanics have evolved significantly, with objects of study becoming complex. While traditional experiments focused on canonical objects like cylinder wake flows and backward step flows, current studies contain complicated cases such as the motion of flying birds (Usherwood et al. 2020). Even studies in canonical objects are now in three-dimensional experiments with perspective effects in the recorded images, which challenge accurate masking. This becomes essential in fluid–structure interaction studies where the object's shape and position must be detected. Current approaches rely on either identifying void regions with no particles (Jux et al. 2021) or case-specific manual masking.

Apart from object identification, flow structures also require segmentation. Turbulent flow structures are known for their complex and chaotic patterns, which make the identification of these structures complicated. The turbulent/non-turbulent interface (TNTI) is among these structures, marking the boundary between the chaotic, rotational regions of turbulent flow and the irrotational regions (Westerweel et al. 2005). Accurately detecting the TNTI structures is essential

for understanding and modelling turbulent properties, such as transport across the interfaces. In scalar turbulence, such as smoke plumes, the TNTI displays a sharp-edge separation from the non-turbulent region. However, edge detection is still challenging due to the chaotic patterns of turbulence (Asadi 2024). Scalar turbulence segmentation is a common measure to detect the TNTI. Existing methods primarily rely on a threshold approach, finding the local minimum from the histogram intensity distribution, or a clustering approach (Younes et al. 2021). Manual segmentation to detect turbulent structures is also employed in complex situations. A review of TNTI detection algorithms is available in the thesis of Asadi (2024). All these examples highlight the need for a universal segmentation model that can effectively work on both coherent structures and objects in fluid experiments.

Recent advancements in image segmentation using artificial intelligence (AI) offer favourable applications in fluid mechanics experiments. Vennemann and Rösgen (2020) introduced an automatic masking method based on artificial neural networks (ANNs) in velocimetry images. This approach showed promising results in 2D measurements, particularly in scenarios where only a single object is present within the field of view. However, its capability to segment complex flow structures or differentiate between distinct objects in the view, such as for segmenting the bike from the cyclist in a sports flow experiment (Jux et al. 2018), is constrained. The Segment Anything Model (SAM), developed by Meta AI (Kirillov et al. 2023), stands out as a

✉ Ali R. Khojasteh  
A.R.Khojasteh@tudelft.nl

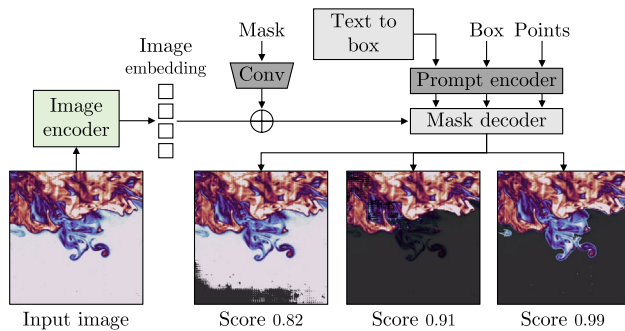
<sup>1</sup> Laboratory for Aero and Hydrodynamics, Delft University of Technology, 2628 CD Delft, The Netherlands

foundation model. The extensive training dataset of SAM, consisting of over 1 billion masks and 11 million images, offers a robust starting point for exploring its capabilities in fluid experiments.

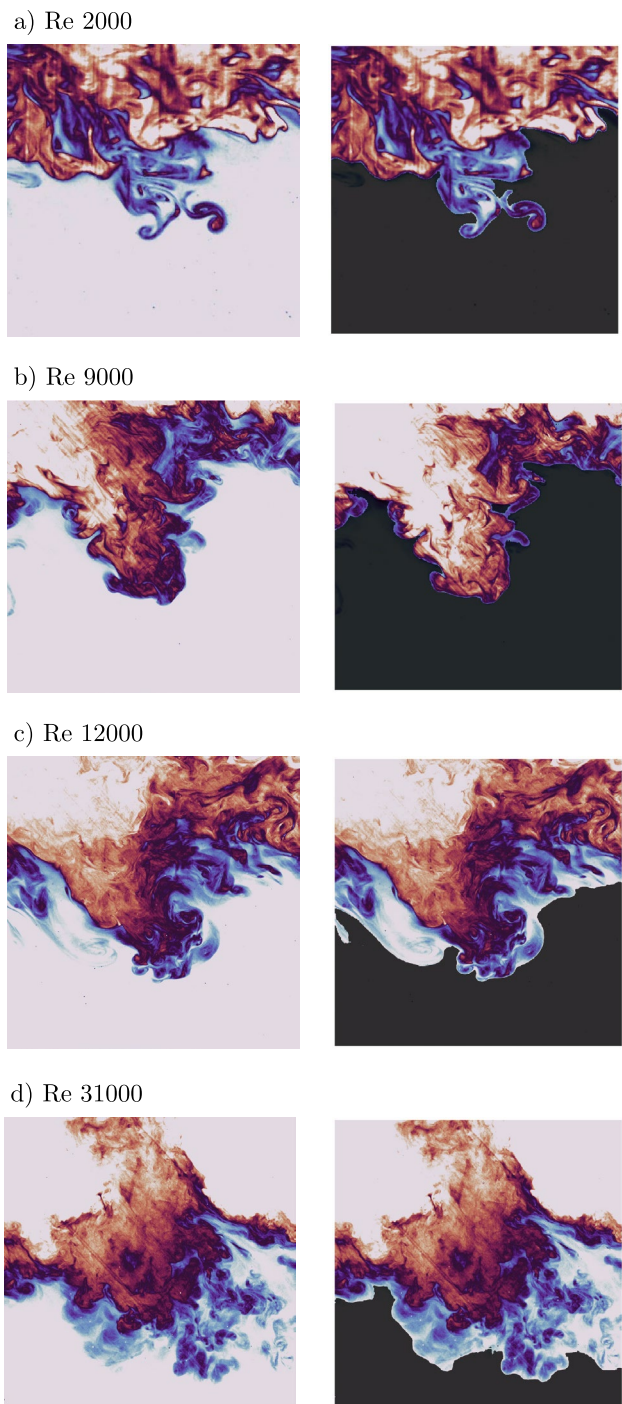
In this paper, we begin by introducing object segmentation and detailing the process of fine-tuning to address the complexities in fluid flow detection using pre-trained architecture and weightings of SAM. We focus on detecting structures in a turbulent flow, the turbulent/non-turbulent interface, by using a time series of images of scalar concentration. This needed fine-tuning of the mask decoder of the SAM implementation. We fine-tuned the mask decoder using the same approach as the model was originally trained, aiming to detect scalar turbulent/non-turbulent structures. Finally, we demonstrate how the prompt encoder of SAM can be modified and combined with language models to ease complex object detection in experiments using only textual input.

## 2 Segment Anything Model

SAM’s architecture comprises three key modules: an image encoder, a prompt encoder, and a mask decoder (see Fig. 1). The image encoder processes input images to generate image embeddings (representations), while the prompt encoder transforms point and box prompts into embeddings that guide segmentation. The mask decoder combines the information from the image and prompts encoders to predict the final segmentation mask(s). SAM accepts guiding prompts in various forms, such as points or bounding boxes. However, complex geometries require more detailed prompts, and objects often move during experiments, such as a flying bird (Usherwood et al. 2020). Therefore, using point or box prompts might not be directly applicable in fluid experiments.

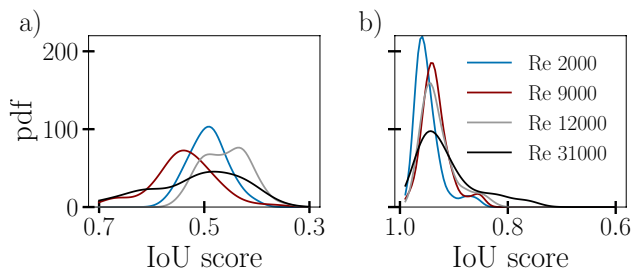


**Fig. 1** The Segment Anything Model (SAM) (Kirillov et al. 2023). Its image encoder creates a representation that supports object mask generation in response to different prompts, capable of producing multiple valid masks (shown in black) and confidence scores



**Fig. 2** Identification of turbulent/non-turbulent flow. The model is trained on **a** a low Reynolds number of 2000. **b–d** The fine-tuned model is then applied at higher Reynolds numbers

Recent studies have attempted to integrate natural language models, such as BERT introduced by Google AI Language (Devlin et al. 2018), as encoder prompts to perform highly specific and context-aware tasks. The language understanding from BERT helps the model focus attention



**Fig. 3** IoU scores for four Reynolds numbers: **a** SAM and **b** Fine-tuned SAM in TNTI detection

and isolates desired objects within an image. In this study, we were inspired by Lightning AI (Lightning 2024), which integrated natural language prompts with GroundingDino (Liu et al. 2023) and SAM. GroundingDino employs BERT to detect a bounding box around objects. BERT tokenises the textual input to create contextualised embeddings, which are enhanced using text-to-image and image-to-text cross-attention mechanisms (Liu et al. 2023). These refined features are processed by a cross-modality decoder, aligning the text with relevant visual regions to generate bounding boxes around described objects and serve as prompts for the SAM model (see Fig. 1). We, therefore, can combine language understanding from BERT to use textual inputs in SAM for flow experiment segmentation.

### 2.1 Fine-tune SAM model

Fine-tuning involves optimising a pre-trained model (architecture+weights) with data specific to a particular use case. Ma et al. (2024) demonstrated that employing SAM in medical images can enhance performance, particularly when the number of training images is substantially increased. The fine-tuning process involves multiple epochs, where the model iterates over the entire dataset, computing the loss between predicted masks and ground truth masks for each batch and updating the model's parameters using backpropagation. During fine-tuning, the pre-trained model's parameters are adjusted to minimise the discrepancy between the predicted segmentation masks and the ground truth masks. This is achieved by iteratively optimising the model's parameters using an optimisation algorithm (in this case, Adaptive Moment Estimation, ADAM (Kingma and Ba 2014)). The loss between the predicted and ground truth penalises deviations between the predicted and ground truth masks. Through this process, the model learns better to capture the specific patterns and features in datasets, ultimately improving its performance on segmentation tasks such as scalar turbulence. The convergence and evaluation of the fine-tuning process are provided in Appendix A.

## 3 Detect scalar turbulence

Direct application of SAM to scalar turbulence works on random cases, but as Fig. 1 also illustrates, the output valid masks fail as they are trained and designed for natural images. SAM suffers from incorrect predictions, broken masks, and large turbulent and non-turbulent detection errors. Scalar turbulence can exhibit complex patterns, low-contrast boundaries, thin structures, and significant differences from the objects typically found in natural images. Despite being trained with 1.1 billion masks, SAM's prediction quality falls short in dealing with turbulent flow.

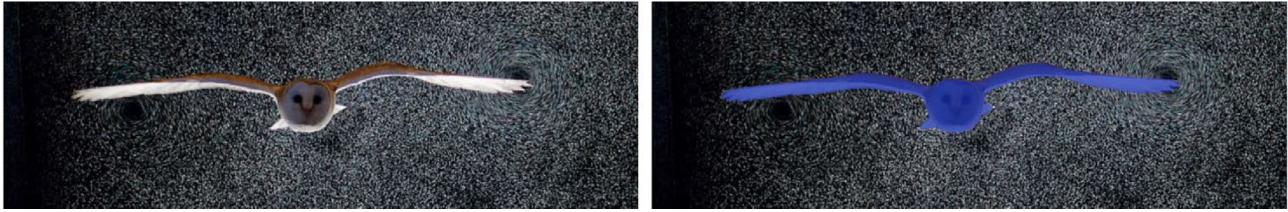
We then fine-tuned the mask decoder of SAM for the specific task of scalar turbulence segmentation. As explained in Appendix A, we selected reliable turbulent/non-turbulent masks from low Reynolds number experimental data and applied them to pre-trained SAM model weights. We used scalar turbulence images of a jet flow provided by Fukushima and Westerweel (2022). We intentionally trained the model with low Reynolds number data because the interface is well-shaped at such Reynolds numbers. Subsequently, we applied the fine-tuned model to higher Reynolds numbers, as shown in Fig. 2 and evaluated in Fig. 3. The performance of the fine-tuned model improved significantly, with IoU scores, which measure the overlap between the predicted and reference masks relative to the area of their union (see Appendix A), increasing from 0.5 to above 0.95 for all Reynolds numbers. At higher Reynolds numbers, the interfaces have more scattered patterns and less sharply defined edges, which is why the pdf plots show broader distributions. We extend an additional application of the segmentation model in bubbly flows in Appendix B.

## 4 Objects in experiments

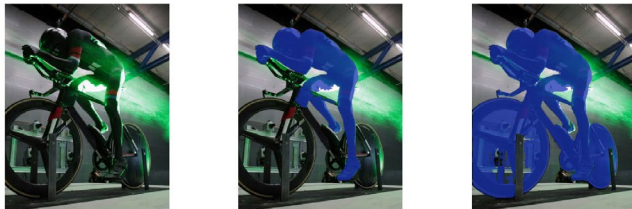
We analysed recent particle image velocimetry (PIV) and particle tracking velocimetry (PTV) experiment images, which present unique challenges compared to other segmentation cases due to thousands of surrounding light-emitting particles. In our initial case study, we focused on volumetric measurements of vortices behind a flying owl as it crossed the field of view (Usherwood et al. 2020). Given the complex nature of this object, manual masking proved impractical. Instead, by inputting only the text "Flying Owl", the segmentation model accurately produced masks without necessitating fine-tuning (see Fig. 4). The next case study involved 3D-PTV analysis around a cyclist (Jux et al. 2018). The model segmented the cyclist and accurately differentiated the bike from the cyclist's body.

The most challenging scenario occurred during particle detection in a water tank experiment where the jet was

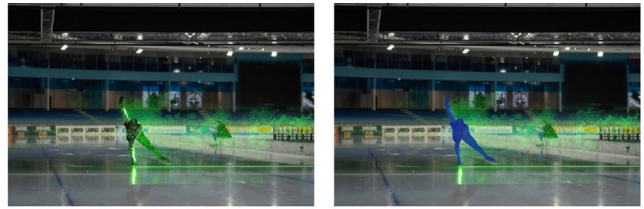
a) Text input: Flying owl



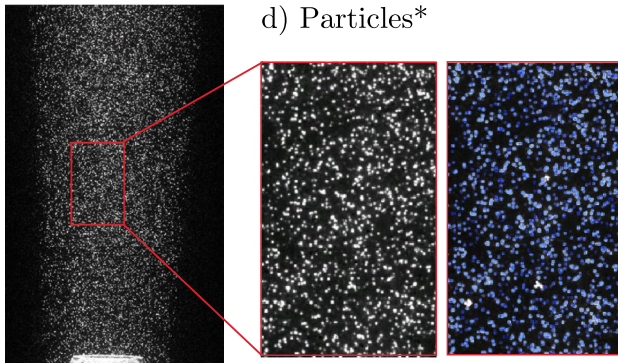
b) Cyclist + bike



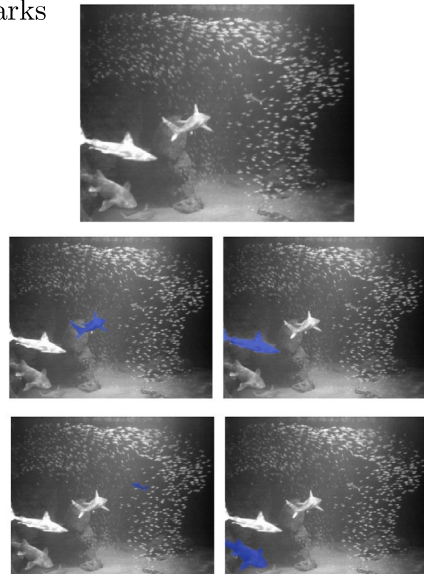
c) Skater



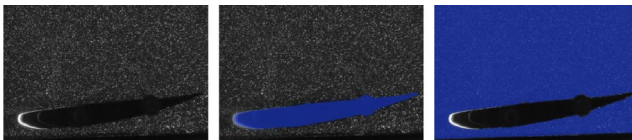
d) Particles\*



e) Sharks



f) Hydrofoil + wall



**Fig. 4** Object detection within the field of view of flow experiments. The segmentation model demonstrates robustness across various objects, accurately detecting their position, area, and shape with solely textual inputs. Textual inputs are **a** A flying owl (Usherwood

et al. 2020), **b** A cyclist and a bike (Jux et al. 2018), **c** A skater (Terra et al. 2023), **d** Particles (Schanz et al. 2016), require fine-tuning similar to Appendix B, **e** Sharks (Muller 2022), **f** A hydrofoil close to a wall (Zhou 2023)

injected into the tank (Schanz et al. 2016). This case required fine-tuning the segmentation model for effective particle detection (similar to the bubbly flow in Appendix B). Even after fine-tuning, some large particles remained undetected. We found the model fails to detect particles when the background is fully dark, and the particles are bright. Therefore, we inverted the image colours to have dark particles and a bright white background. Furthermore, the model effectively masked four sharks in a 3D-PIV study of schooling fish (Muller 2022). This level of understanding from the model to distinguish sharks individually allows for tracking

schooling fish without interference from larger fish. In 2D-PIV image analysis, the model segmented a flat plate and a hydrofoil using the "Hydrofoil + Wall" input (Zhou 2023).

## 5 Conclusion

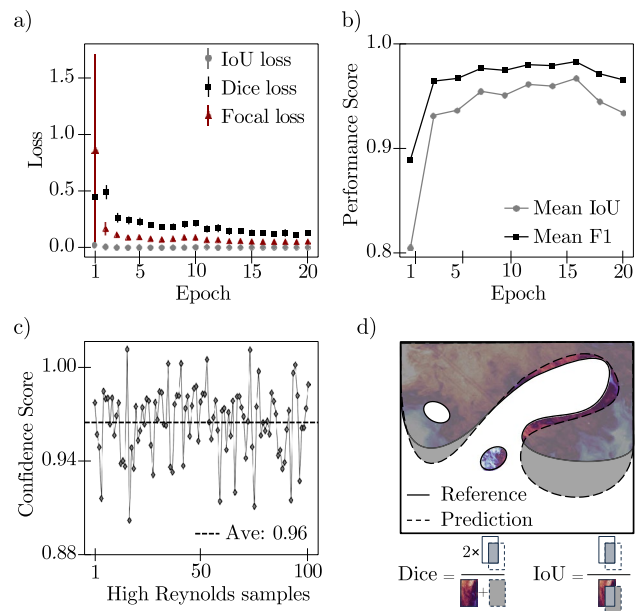
In conclusion, we have introduced a practical approach by implementing the natural image segmentation model for coherent structure and object identification in fluid experiments. SAM has proven to be a valuable tool, capable of

being fine-tuned to understand the TNTI structures. Our approach involved fine-tuning the mask decoder of the SAM model, aligning with its original training methodology. Additionally, we integrated the model with language models to serve as a prompt encoder, allowing communication between the language model and SAM for precise context detection.

## Appendix A: Fine-tuning and optimisation using adaptive moment estimation

SAM used the Adaptive Moment Estimation (ADAM, Kingma and Ba 2014) optimisation algorithm during the training of its model. We implement a similar optimiser for our fine-tuning process. The fine-tuning was performed on A100 NVIDIA GPUs. The key idea behind ADAM is to maintain an exponentially decaying average of past and squared gradients and then use these averages to update the parameters. To quantify the performance of the turbulent/non-turbulent interface (TNTI) identification, we used three loss functions: Focal loss, Dice loss, and Intersection over Union (IoU) loss as explained by Kirillov et al. (2023). Focal loss addresses class imbalance by down-weighting easy examples and focusing more on hard examples. The Dice loss compares the overlap between the predicted and ground truth masks across the entire area. In contrast, the IoU measures the overlap relative to the area of their union, as illustrated in Fig. 5d. The ground truth (or reference) is the interface obtained from conventional edge detection techniques from Westerweel et al. (2005), which was first introduced by Prasad and Sreenivasan (1989). The total loss is then computed as the sum of the weighted Focal loss (with a weight of 20), Dice loss (with a weight of 1), and IoU loss (also with a weight of 1), reproducing the training configuration of the original SAM model.

We divided our data into 70% of uncorrelated snapshots (random snapshots) to train the model and kept the rest of 30% to evaluate. This means that the model did not have access to the 30% of data. During training, losses converged after ten epochs, as illustrated in Fig. 5a. We validated the model every two epochs using the remaining 30% of images unseen by the model, as shown in Fig. 5b. Validation scores ranged between 0 and 1, where 1 indicated perfect segmentation with respect to the ground truth. Both Dice coefficient (F1) and mean IoU scores remained well above 0.97 after epoch one but began to decline after 15 epochs, indicating overfitting. Based on this analysis, we stopped fine-tuning at epoch 15. We then applied the fine-tuned model to cases with higher Reynolds numbers. As shown in Fig. 5c, the predicted IoU scores average

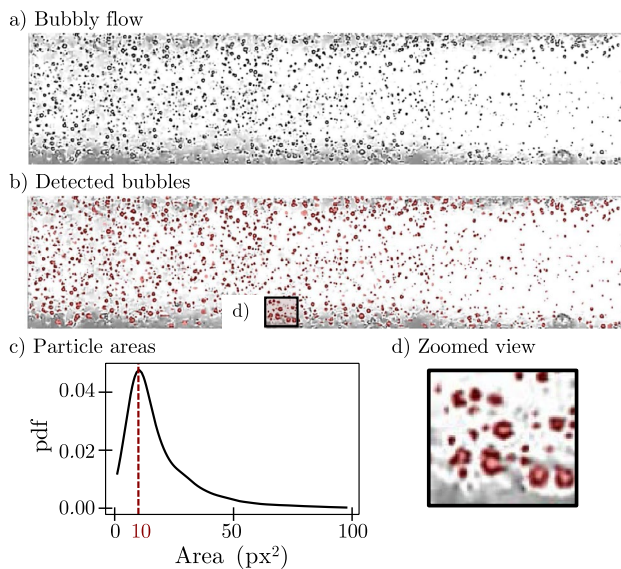


**Fig. 5** Performance and loss metrics over 20 epochs during the training of the segmentation model on TNTI images. **a** Displays the average IoU loss, Dice loss, and Focal loss for each epoch, with error bars indicating the standard deviation. **b** is the Mean IoU and Mean F1 scores over unseen 30% samples of low Reynolds number. **c** Shows the confidence and IoU predictions on high Reynolds cases (see Fig. 2). **d** Schematic of IoU and Dice losses for TNTI detection

approximately 0.95. This high average score indicates that the model predicts the results with a high degree of confidence. All technical steps are available in the repository ([https://github.com/AliRKhojasteh/Flow\\_segmentation](https://github.com/AliRKhojasteh/Flow_segmentation)).

## Appendix B: Bubbly flows

Here, we focused on the hydrodynamics of bubbles in a channel (Hreiz et al. 2015). In this scenario, bubbles travel along the channel and experience size changes over time. The challenge arises from the fact that bubbles, while not conventional objects have hydrodynamic properties absent from the SAM model's trained dataset. Moreover, bubbles exhibit diverse sizes and shapes, further complicating their detection. Therefore, fine-tuning the SAM model becomes necessary. To achieve this, we employed manual masks to train the model on bubble shapes and sizes. Consequently, the model successfully identified a wide range of bubbles, spanning from dot shapes to fully formed bubbles. The distribution of bubble areas in the probability density function (PDF) reveals the entire spectrum of sizes, with an average diameter of 3.5 pixels (see Fig. 6).



**Fig. 6** **a** Illustration of bubble detection in a complex bubbly flow (Hreiz et al. 2015). **b–d** The fine-tuned SAM successfully identifies approximately 3000 bubbles, handling a variety of shapes and sizes

**Author contributions** ARK wrote the main manuscript. WvW and JW contributed to the project conceptualisation with the corresponding author. All authors reviewed the manuscript.

**Data availability** To facilitate the application, we have established a repository, [https://github.com/AlirKhojasteh/Flow\\_segmentation](https://github.com/AlirKhojasteh/Flow_segmentation), where models and usage examples can be accessed.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Asadi M (2024) Exploring Turbulence - Turbulence Interactions: Impacts of Incoming Turbulence on Wall-Bounded Flows. Ph.D. thesis, Norwegian University of Science and Technology. <https://hdl.handle.net/11250/3115493>
- Devlin J, Chang M-W, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding, [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
- Fukushima C, Westerweel J (2022) Original data for the combined PIV/LIF measurement of a turbulent jet at a Reynolds number of 2000. <https://doi.org/10.4121/14226458.v2>
- Hreiz R, Abdelouahed L, Fünfschilling D, Lopicque F (2015) Electro-generated bubbles induced convection in narrow vertical cells: PIV measurements and Euler–Lagrange CFD simulation. *Chem Eng Sci* 134:138. <https://doi.org/10.1016/J.CES.2015.04.041>
- Jux C, Sciacchitano A, Schneiders JF, Scarano F (2018) Robotic volumetric PIV of a full-scale cyclist. *Exp Fluids* 59:1. <https://doi.org/10.1007/s00348-018-2524-1>
- Jux C, Sciacchitano A, Scarano F (2021) Object surface reconstruction from flow tracers. *Exp Fluids* 62:42. <https://doi.org/10.1007/s00348-021-03139-1>
- Kingma DP, Ba J (2014) Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
- Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC et al. (2023) Segment anything. [arXiv:2304.02643](https://arxiv.org/abs/2304.02643)
- Lightning AI, PyTorch Lightning Onboarding, <https://lightning.ai/onboarding>, accessed: February 15, 2024
- Liu S, Zeng Z, Ren T, Li F, Zhang H, Yang J, Li C, Yang J, Su H, Zhu J et al. (2023) Grounding dino: Marrying dino with grounded pre-training for open-set object detection, [arXiv:2303.05499](https://arxiv.org/abs/2303.05499)
- Ma J, He Y, Li F, Han L, You C, Wang B (2024) Segment anything in medical images. *Nat Commun* 15:1. <https://doi.org/10.1038/s41467-024-44824-z>
- Muller K (2022) Tracking Schooling Fish in Three Dimensions: Experiments at the Rotterdam Zoo. Ph.D. thesis, Delft University of Technology. <https://doi.org/10.4233/uuid:12905cc1-b9e7-4bff-972d-19bea5cf4fdf>
- Prasad RR, Sreenivasan KR (1989) Scalar interfaces in digital images of turbulent flows. *Exp Fluids* 7:259. <https://doi.org/10.1007/BF00198005>
- Schanz D, Gesemann S, Schröder A (2016) Shake-The-Box: Lagrangian particle tracking at high particle image densities. *Exp Fluids* 57:1. <https://doi.org/10.1007/s00348-016-2157-1>
- Terra W, Spoelstra A, Sciacchitano A (2023) Aerodynamic benefits of drafting in speed skating: estimates from in-field skater's wakes and wind tunnel measurements. *J Wind Eng Ind Aerodyn* 233:105329. <https://doi.org/10.1016/j.jweia.2023.105329>
- Usherwood JR, Cheney JA, Song J, Windsor SP, Stevenson JPJ, Dierksheide U, Nila A, Bomphrey RJ (2020) High aerodynamic lift from the tail reduces drag in gliding raptors. *J Exp Biol* 223:214809. <https://doi.org/10.1242/jeb.214809>
- Vennemann B, Rösgen T (2020) A dynamic masking technique for particle image velocimetry using convolutional autoencoders. *Exp Fluids* 61:168. <https://doi.org/10.1007/s00348-020-02984-w>
- Westerweel J, Fukushima C, Pedersen JM, Hunt JCR (2005) Mechanics of the turbulent-nonturbulent interface of a jet. *Phys Rev Lett* 95:174501. <https://doi.org/10.1103/PhysRevLett.95.174501>
- Younes K, Gibeau B, Ghaemi S, Hickey JP (2021) A fuzzy cluster method for turbulent/non-turbulent interface detection. *Exp Fluids* 62:1. <https://doi.org/10.1007/S00348-021-03169-9>
- Zhou S (2023) Lift Coefficient of an Accelerating Wing with Ground Effect, master thesis, Delft University of Technology. <http://resolver.tudelft.nl/uuid:b7301745-e503-4a76-83d9-aea7d2c19437>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.