MACHINE LEARNING
Science and Technology

PURPOSE-LED PUBLISHING™

Part of the MACHINE LEARNING Series

BENCHMARK • OPEN ACCESS

Domain-randomised instance-segmentation benchmark for soot in PIV images

To cite this article: Basil Jose *et al* 2025 *Mach. Learn.: Sci. Technol.* **6** 040504

IOP Publishing Mach. Learn.: Sci. Technol. 6 (2025) 040504 https://doi.org/10.1088/2632-2153/ae2565

MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
1 August 2025

REVISED
8 November 2025

ACCEPTED FOR PUBLICATION
27 November 2025

PUBLISHED
15 December 2025

BENCHMARK

Domain-randomised instance-segmentation benchmark for soot in PIV images

Basil Jose[1], **Klaus Peter Geigle**[2] **and Fabian Hampp**[1,*]

[1] Institute of Combustion Technology for Aerospace Engineering, University of Stuttgart, Pfaffenwaldring 38-40, 70569 Stuttgart, Germany
[2] German Aerospace Center (DLR), Institute of Combustion Technology, Pfaffenwaldring 38-40, 70569 Stuttgart, Germany
[*] Author to whom any correspondence should be addressed.

E-mail: fabian.hampp@ivlr.uni-stuttgart.de

Keywords: optical diagnostics, instance segmentation, PIV, image segmentation, domain randomization, sooting flame, RQL combustor

Supplementary material for this article is available online

Abstract

Access to high-level statistical information from scientific image-based diagnostics is facilitated by the capacity to segment quantity-of-interest (QoI) signal accurately from measurement noise and interferences. In particular under conditions with diverse nuisance signatures, deep learning (DL) pipelines can outperform algorithm-based computer vision (CV) strategies and simultaneously enhance domain-invariance. DL-pipelines aim to overcome the generalisation bottleneck through augmentations and learnable parameters, supplemented by huge amounts of data. Yet, manual pixel-accurate annotation for scientific tasks is prohibitively expensive and the robustness of the trained models is often impeded by the underlying narrow and domain-specific set of training data. In the current study, a previously developed domain-randomised pipeline for automatic annotation and synthetic training data generation is benchmarked using a physics-aware composite score. An ablation study for background variance, QoI placement strategy and QoI object source proportions is conducted. The benchmark, consisting of 72 synthetic training data generation recipes, highlights the optimal set of domain-randomisation parameters to strike the balance between domain-invariance and segmentation accuracy. Synthetic training datasets with a balanced mix of QoI objects, fused onto realistic background instances, are found to provide the most accurate models for segmentation. The best model is subsequently used for the inference on semantically challenging Mie scattering images containing particle-image velocimetry (PIV) tracer particles and soot filaments. The model's ability to detect soot structures with good accuracy is demonstrated and high-level soot filament area and contour statistics (e.g. curvature and fractal dimension) are used to delineate the effects of turbulent flow on soot filament structures. The present study highlights key parameters to tune domain randomisation strategies for DL-training pipelines and the pipeline usability and transferability is proved for autonomous semi-supervised learning. This eases access to high-level statistics in scientific image-based diagnostics.

1. Introduction

Across virtually every branch of experimental science and engineering, the extraction of physically high-level meaningful information requires isolating a structured signal from layers of stochastic noise, systematic distortions and instrument-specific artefacts. Optical measurements, such as flow-diagnostic techniques, including planar laser-induced fluorescence (PLIF), particle-image velocimetry (PIV), laser-induced incandescence and Schlieren imaging often operate on relatively weak photon fluxes that are challenging to be separated from multiple spurious sources [1, 2]. In simultaneous multi-scalar and flow field measurements, unambiguous signal detection and classification remain non-trivial, while sophisticated and computationally expensive post-processing techniques are being developed [3–6]. A successful quantity-of-interest (QoI) segregation from background and measurement nuisance facilitates a quantitative description of, e.g. scalar boundary statistics or, in combination with flow field measurements, turbulence–chemistry interactions. Such measurements are indispensable for a wide range of turbulent reacting flows and validation of corresponding numerical models [4, 7, 8]. Shot and camera read-out noise erode the signal-to-noise, while non-uniform laser-sheets and beam-steering by refractive-index gradients introduce large-scale intensity/separation bias and noise [9–11]. In general, physical signal separation (temporal or spectral) is always preferred. For example, in simultaneous PIV / PLIF measurements, wavelength-based segregation of the resulting signal by means of dichroic beam splitters onto multiple cameras is effective to conduct simultaneous multi-scalar and flow field measurements [12, 13]. However, in elastic light scattering, extinction, shadowgraphy, or Schlieren imaging, physical segregation techniques are predominantly not feasible. For example, PIV, on the basis of flow-tracer Mie scattering, delivers high-fidelity flow-field measurements across applications ranging from canonical turbulent channel-flow experiments [14], swirl-stabilised combustor diagnostics [15] and aero-acoustic tests [16]. In multi-phase flows, the Mie scattering from PIV tracer particles overlaps spectrally and spatially with secondary scatterers (e.g. spray droplets or soot filaments), complicating downstream processing [15, 17]. The same holds for shadowgraphy imaging, where an accurate separation of signal sources is critical [18, 19]. In the aforementioned examples, the signal source separation must be conducted by computer vision (CV) based post-processing steps, yet can offer access to high-level statistics at minimum hardware cost. Conducting such measurements in confined setups with limited optical access further complicates the diagnostic setup and the required post-processing steps due to the rise of nuisances [20, 21]. The latter include window fouling, stray reflections, pixel saturation, vignetting and sensor fixed-pattern noise, compounding in segmentation challenges of the true signal [22–25]. These must be either pre-compensated in hardware or post-hoc corrected by sophisticated calibration and deconvolution routines [26, 27]. Once more, physics-based correction methods, e.g. light sheet intensity or *in-situ* background corrections, are preferred and generally can enhance any downstream segmentation and classification accuracy. Generally, the detection and extraction of QoI from multi-scalar scattering images is hampered by: (a) Optical semblance of scalar-scattering (signal), measurement nuisance (scattering-interference), and accompanying background (noise); (b) Presence of severe saturation effects due to variance in scatterer size, e.g. PIV tracer particles $\sim$1 μm, droplets 10–100 μm, wall 1–100 mm, and shot noise [18]. The aforementioned challenges can lead to detection inaccuracies that distort the statistics needed to understand turbulence-scalar coupling, hindering corresponding technology development, e.g. injector/combustor design and optimisation for emissions control and cleaner combustion. Classical threshold- or morphology-based pipelines often fail when illumination or other domain conditions drift from the calibration [28–30]. Robust quantitative analysis therefore demands either extensive manual tuning or data-driven models that can learn to discount nuisance variability while preserving the spatial resolution of signal features.

Recently, data driven and deep learning (DL) based approaches, with their ability to handle complex tasks, have proved effective in diverse domains including natural language processing [31], robotic systems [32], weather forecasting [33], finance and banking [34], and fluid dynamics research [35, 36]. In optical diagnostics, such methods have already been successfully applied to tasks such as fuel droplet and gas liquid interface analysis in shadowgraphy [21, 37], PLIF [38], flame front extraction in OH-PLIF [30], and light-sheet fluorescent imaging [39], amongst others. This can be attributed to the generalisation capabilities of DL-models that circumvent the constrains of vast image content diversity in these experiments. Yet, two key obstacles remain: (a) lack of labelled data as annotating pixel-accurate masks for thousands of megapixel frames demands prohibitive expert input [40]. (b) Domain-shift, where networks trained on broad natural-image corpora (e.g. MS-COCO [40]) do not transfer to the specific harsh illumination, noise statistics, and filamentary structures encountered in scientific diagnostics. Computer vision models for image segmentation require carefully curated masks for QoI to create the training data. The immense effort of manual labelling can be circumvented to some extent by the use of automated synthetic labelling strategies such as teacher-student pseudo-labelling frameworks [41] and zero-shot SAM-style pre-labelling [42, 43]. Bayesian uncertainty engines can quantify mask reliability and guide human correction, e.g. BNN phase-imaging [44] or Bayesian SR-SIM in microscopy [45], but again assume some ground truth. These methods also require initial manual calibration with limited context awareness and robustness [46, 47]. Domain randomisation (DR), a novel method first applied to transfer domains in robotics [48] and later in many CV applications [21, 49, 50], can effectively generalise to a wide range of domains, while accurately segmenting QoI with minimal human input. While this reduces the annotation effort considerably with refined control over the training data aesthetics, effective domain randomisation demands careful calibration of its parameters (augmentation strength and type, background variations, QoI object-mix ratios, class selection, fusion methodologies) for robust generalisation without too much noise that can overwhelm the model [51]. Excessive randomness diminishes fidelity, while too little induces over-fitting to synthetic quirks hindering robustness. Despite its central importance, the community lacks systematic guidelines for tuning DR hyper-parameters, leading to poor reproducibility and interpretability accommodated with significant time and computational resource requirements for isolated experiments to achieve the best fit. While adaptive, active, or Bayesian DR techniques solve this by automatically learning the statistical domain distributions, they remain a black box, require extensive computational effort, and depend heavily on a differentiable renderer [51–54]. These constraints may be the reason for hindrance of wider exploitation of this technique to other CV tasks such as instance segmentation and highly relevant fields of study such as optical diagnostics in fluid mechanics and combustion science which often are challenging for conventional methods.

An in-house DL-augmented data-analysis framework based on domain randomisation is expanded from droplet and spray detection in shadowgraphy images [21, 37], to robust, instance-level segmentation of soot filaments in elastic light scattering / PIV measurement in a swirl model combustor [15, 55]. For the dataset used in the present work, it is not trivial to segregate (or annotate) manually between the in- and out-of-plane soot filaments (QoI signal and nuisance), scattering from PIV tracer particles (nuisance), and background (noise) due to their similarities in intensities and low dynamic range. The efficacy of the in-house automatic synthetic data generator is validated in generating 72 training data recipes, each having $\sim$10k pixel-level annotated images and are varied by three background families (real/COCO/black), three placement policies (blend/replace/random), and five object-class proportions. For the subsequent comprehensive parameter study, a new dimensionless metric that weights soot priority, refined masks, size stratification and failure modes tailored to filamentary soot and PIV nuisances is introduced to quantify the underlying interacting factors for better model performance. A further physical model validation is done by conditional statistics on the model inference ($\sim$24k raw images). Thus, for scientific imaging, where pixel-accurate QoI segmentation is non-trivial with the risk of domain drift (even for validation data), the current study makes the following contributions:

- Purely synthetic, domain-randomised DL-pipeline for instance segmentation in elastic light scattering (ELS) images of turbulent, reacting multiphase flows with PIV tracers and soot filaments.
- 72-recipe benchmark and ablation spanning object mix, background families, and placement/fusion policies for DR datasets.
- Physics-aware, soot-centric composite evaluation metric addressing the inadequacy of standard DL-scores for high-accuracy scientific imaging.
- Physical validation on 24.6k ELS images with advanced conditional soot-structure statistics for flow/-combustion mapping.
- Reproducible DR setup with practical guidelines for generating tunable synthetic datasets for combustion imaging.

In the next section, the experimental and diagnostic setups are briefly introduced, followed by detailed description of the methodology of the DL-approach, benchmarking metrics in section 3. The sensitivity of training and model parameters is discussed, including their validation, in detail in section 4 that is followed by the application of the best model to the current experimental dataset to extract physical insights from the sooting flames.

2. Experimental setup

The description of the combustor and optical setup widely follow the existing primary literature of the case on how the data used here are generated. To ease understanding of the experimental conditions, relevant parts of the experimental description are provided in the following.

2.1. Aero-engine model combustor

The dataset has been acquired in the single-sector, swirl-stabilised RQL-model combustor described in detail by Stöhr *et al* [6] and references therein. The burner geometry is comprised of two concentric, swirl air nozzles with inner diameter 12.3 mm and outer diameter 19.8 mm. Gaseous ethylene is injected through 60 channels ($0.5 \times 0.4\,\mathrm{mm}^2$) located between the two air streams, ensuring partial-premixing prior to ignition. The nozzle exits are mounted flush with the burner base plate of a square combustion chamber ($68 \times 68\,\mathrm{mm}^2$, height 120 mm) that is fitted with 3 mm fused-silica quartz glass windows on all four sides to provide full optical access. A cylindrical exhaust duct ($d = 40$ mm) removes the products downstream into the pressure housing. Four secondary-air ports ($d = 5$ mm), located 80 mm above the combustor nozzle exit plane, allow independent control of the dilution air and prove critical for the rich/lean pocket dynamics as reported in [15, 17, 56]. The combustor was operated at a moderate pressure of $P = 3$ bar with an overall thermal power of 38.6 kW. The global equivalence ratio, including secondary air, was $\Phi_{\text{global}} = 0.86$, whereas the primary injector was operated at $\Phi_{\text{primary}} = 1.20$. This results in a partially premixed, turbulent swirl flame, whose inner recirculation zone (IRZ) intermittently fills with either lean burned gas or fuel-rich burned gas with low strain and residence time that promotes soot formation [6].

2.2. Optical diagnostics and image acquisition

A high-speed stereoscopic PIV system, combined with planar OH laser-induced fluorescence (OH-PLIF), provided the time-resolved datasets used in this work [15]. Illumination for PIV was provided by a dual-cavity, diode-pumped solid-state laser (EdgeWave IS200-2-LD) emitting 9 mJ pulses at 532 nm and 9.3 kHz repetition rate. The laser beams were shaped via cylindrical lenses into a light sheet with 40 mm height, traversing vertically through the chamber's central plane. The flow was seeded with TiO_2 particles ($d_{\text{p}} \approx 1\,\mu$m) for the PIV flow field measurements. The soot filaments, formed in the post-flame zone, scatter light at the same wavelength as the PIV tracer particles (i.e. 532 nm) and form the QoI for the present study. Two LaVision HSS8 CMOS cameras ($512 \times 640\,\text{px}^2$) recorded the stereoscopic PIV image pairs with inter-frame time of $10\,\mu$s. For OH-PLIF, excitation with a pulse energy of $226\,\mu$J was provided by a frequency-doubled dye laser (Sirah Cobra-Stretch HRR) pumped by a frequency-doubled DPSS laser (EdgeWave IS-811E). A LaVision high-speed intensifier relay optics unit and HSS6 camera, coupled with 100 mm f/2.8 UV lens and a band-pass filter, captured OH-PLIF at 3.1 kHz. The PLIF images were timed between the two PIV laser pulses of every third PIV recording. A total of 24 576 flow vector fields and 8192 OH scalar fields were captured. Velocity evaluation employed 48×48 px interrogation windows with 50% overlap, whereas all segmentation experiments operated directly on the raw sensor resolution. Only the elastic light scattering PIV frames, with both tracer-particle and soot scattering, enter the current segmentation benchmark. For further details on the experimental or diagnostic setup, the reader is referred to Geigle *et al* [17, 56]. The conventional way of extracting soot filaments from the PIV frames is described in [15]. Although effective and statistically accurate, such methods are domain-sensitive, and can selectively cause misclassification and false positives (FPs) that have to be avoided. Examples of these limitations of the conventional method are demonstrated in the supplementary material.

3. Methodology

3.1. Domain randomisation for soot detection

The method is applied in detecting soot filaments from ELS images of PIV measurements in reacting multiphase flows. The Mie scattering from the objects in the laser sheet path is captured using a camera, producing a high-resolution image as shown in figure 1 (left). The background is predominantly dark with the high intensity regions mapping to soot filaments and the gaseous reactant stream mixed with tracer particles (seeding). Due to thermal expansion, the particle seeding density in the combustion products is significantly lower [4], giving rise to the different aesthetic appearance of the seeding material. The low intensity soot filaments are objects that are out of the focal plane of the laser. For automatic mask generation for the QoI, the methodology of Jose *et al* [37] is adopted. For this purpose, a database of soot, out-of-plane regions (OOPObj) and high density and intensity PIV tracer particle regions is generated (see figure 2). To generate synthetic training data, these QoI image snippets are randomly sampled and placed with random translation into random background domains that are assembled from a database of backgrounds, see figure 3. The combined geometric and photometric augmentations with random backgrounds inject enough domain variance for the model to learn invariance to nuisance background during training. Each placed object yields an exact binary mask. The pairs are then converted to COCO style annotations for training, thus providing ground truth information. Iteratively, a complete dataset with train, test, and validation splits is generated.

3.1.1. Database-assembly workflow

A comprehensive database having diversified QoIs and background elements is compiled to aid the generation of synthetic training data. The methodology comprises a two-step process: first, extracting objects and background patches from their original domains, and subsequently, post-processing these elements to prepare them for synthetic-image assembly. In this study, QoIs and backgrounds from three distinct sources are employed to ensure broad domain coverage and foster model generalisation.

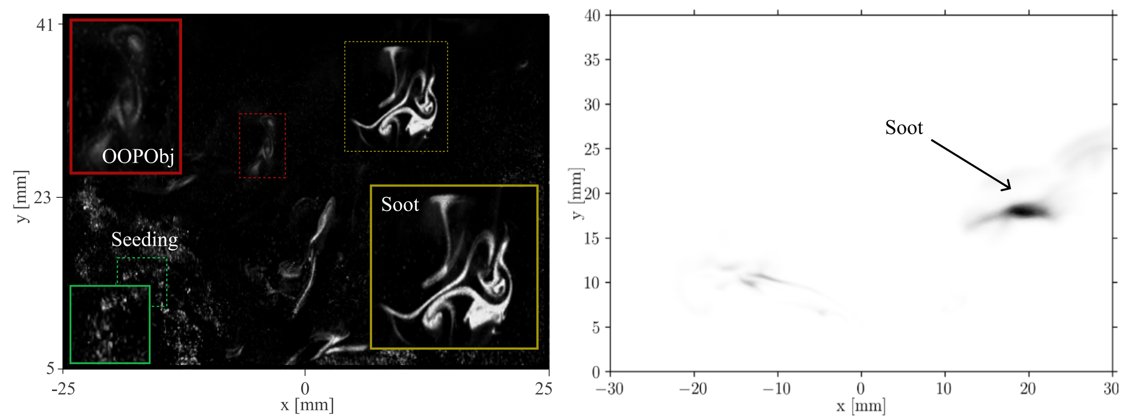

Figure 1. Raw image from PIV Mie scattering (left) and simulated image of soot (right).

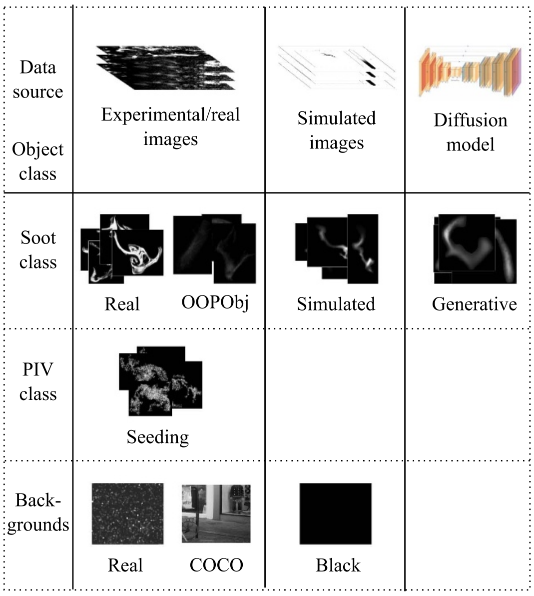

Figure 2. Database generation: The QoIs (soot, OOPObjs, PIV regions) are extracted from the experimental data, simulated images and synthetically generated using diffusion models. They are then classified into two classes, soot and piv along with the background.

Experimental raw data This class of QoIs is meticulously extracted from the raw data preserving their real-scene aesthetics. As the soot and tracer particles scattering share similar greyscale intensity levels and are interlaced by multiphase boundaries, the extraction itself is complex. The following steps are implemented to extract objects-of-interests from the raw data:

- An OpenCV-based connected-components pipeline is applied on each frame to isolate contiguous scatter regions.
- The obtained components are then distinguished into soot filaments in the focus plane of the laser sheet, those out of the laser sheet plane (OOPObjs) and high intensity seeding regions using a combination of their aesthetic properties as well as with pre-existing coarse masks.
- The differentiated objects, boundary padded to mitigate contour detection errors, are stored.
- Considering the vast aesthetic semblance of the QoIs, and therefore the limitations of conventional segmentation methods, a final quick manual check is also done to ensure database classification accuracy and quality control.

CFD/large-eddy simulation (LES) soot fields Soot structures are also extracted from corresponding LES generated soot-fields of the RQL swirl ethylene flame [57]. The simulated scalar fields are inherently without measurement nuisance or other aesthetic features, making the automatic QoI segmentation

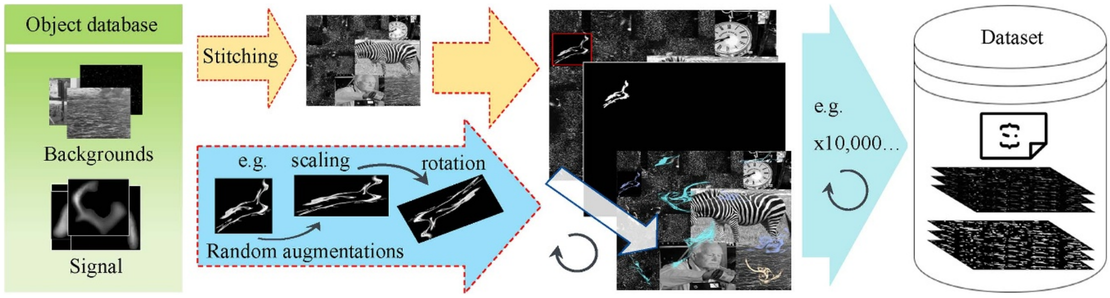

Figure 3. Physics-informed training data generator for automatically generating synthetic images with pixel-accurate annotations for training.

much easier and reliable. Numerical "soot filaments" are detected, extracted, and labelled using the same OpenCV algorithm and stored in a separate class for downstream dataset generation. The simulation images exhibit an inverted greyscale structure, with the dark soot structures shown on a bright background, see figure 1(right). To extract the soot structures in a compatible format, first the images are inverted, artefacts (e.g. axis labels) are removed, followed by the OpenCV morphology analysis to segregate individual soot structures.

Diffusion models A third class of objects is produced with a denoising diffusion probabilistic model [58] trained on the simulated soot filaments. A lightweight U-Net [59] is trained to generate 256×256 single-channel patches of previously unseen soot-like filaments from random noise. A detailed workflow is provided in the supplementary material. Each synthetic greyscale sample can be easily segmented using conventional thresholding to obtain a binary mask that serves directly as ground-truth annotation, eliminating manual labelling. The stochastic generation expands shape, curvature and intensity diversity of the soot structures that are under-represented in the experimental corpus, providing a virtually boundless, diverse supply of realistic QoIs for domain-randomised training.

3.1.2. Background database

A diverse background corpus is essential to prevent the network from over-fitting to spurious cues and to promote robust generalisation. Three complementary background classes are curated and stored as single-channel `uint8` patches:

(a) **In-scene (real) backgrounds.** Patches are harvested from raw experimental images by first removing (zeroing the pixels) soot filaments and bright tracer particle cluster regions with coarse, morphology-based masks. A random seed point is then drawn, a bounding box of random aspect ratio is proposed, and the enclosed region is accepted only if it contains no non-zero pixels, i.e. no soot or high intensity tracer particle regions present. Otherwise a new proposal is generated. Accepted patches preserve authentic sensor noise and optical artefacts of the measurement system.
(b) **Natural-scene (COCO) backgrounds.** To maximise domain shift, random greyscale patches from the COCO dataset [40] are cropped out. These images bear no similarity to the experimental setting and therefore compel the model to discount irrelevant semantic content.
(c) **Uniform black backgrounds.** Zero-valued arrays are generated with OpenCV to provide the degenerate case of minimal visual context and serve as a regulator for evaluation.

During synthetic-image construction, the selected patches are tiled randomly to create a full-resolution canvas, onto which the QoIs are positioned together with their ground-truth masks. This three-pronged strategy widens background variance while keeping the database lightweight and fully automated.

3.2. Dataset generation

A custom Python engine, `SynImgGen`, is developed that assembles fully annotated soot-PIV images automatically using domain randomisation. It combines augmentations (photometric and geometric), multiple classes (for multi-object segmentation), QoI objects (providing semantic diversity) and background variations (for domain-invariance) to generate synthetic images with QoI placed in random background domains. The increased randomness enhances model generalisation to data that is not labelled. Each training image is created on-the-fly by sampling (i) background patches, (ii) a variable-length list of quantities-of-interest, and (iii) a stochastic set of photometric and geometric augmentations. Five QoI types are supported: experimentally extracted soot filaments and OOPObjs, numerically simulated soot contours, diffusion generated soot, and high intensity PIV tracer particle regions. These are classified into two distinct semantic classes, *soot* and *piv* as shown in figure 2. Background samples are drawn with equal probability from either (a) clean in-scene areas of raw PIV frames, (b) greyscale crops of the COCO dataset to induce maximal domain shift, or (c) uniform black canvases that regularise the network.

For every QoI, the generator applies random scaling, rotation, intensity re-weighting, and one of two placement policies: *blend*, which preserves local background luminance to mimic semi-transparent boundaries around soot filament boundary, or *replace*, which substitutes the underlying background pixels with the QoI patch. Overlap between instances is limited by a configurable intersection test. Class proportions are enforced at the image level so that each synthetic frame reflects the global pre-defined target distributions. Corresponding binary masks are rendered at generation time and converted to COCO-style polygon annotations. Datasets are produced in three splits of 6000 training, 3000 testing, and 1000 validation images at predefined spatial resolutions. All metadata (generation parameters, split indices, background provenance) are logged to the dataset to guarantee transparency and reproducibility. The resulting datasets form the basis for all subsequent training and ablation studies.

3.3. Instance segmentation model

Instance segmentation seeks to assign a class label and a pixel-accurate mask to every object, while distinguishing multiple instances of the same class. Unlike semantic segmentation, which merges all pixels or object detection of a class, yielding localised objects with bounding boxes. Instance segmentation combines both capabilities and is therefore well suited to the delicate task of delineating soot filaments midst dense Mie scattering in PIV images. It further facilitates conditional statistics on each individual detected object, thereby offering high-level quantitative information of the underlying physical processes. Mask R-CNN architecture based on Detectron2 framework [60] is used and initialised from MS-COCO weights [61].

Mask R-CNN The canonical Mask R-CNN couples a region-proposal network to a ResNet-50-FPN backbone and augments the detection branch with a four-layer convolutional mask head producing a 28×28 binary mask per ROI [61]. Training is done on the Ehler's computing cluster with a batch size of 16 (NVIDIA A100 GPUs), stochastic gradient descent with momentum 0.9, an initial learning rate of 0.02, and 500-iter warm-up followed by multi-step decays, as per the Detectron2 best practice. Training is done for <270k iterations with metrics validation every 1000 steps.

3.4. Benchmarking

3.4.1. Evaluation metrics

To evaluate the models performance in terms of localisation, class-delineation in soot filaments and tracer-particle regions, the average-precision based methodology used in the COCO evaluation framework [40] is adopted. This combines metrics for object detection (bounding boxes) and instance segmentation (pixel-accurate masks) into a single, well-understood suite. First, for any predicted mask $\mathcal{P}$ and its ground-truth mask $\mathcal{G}$, the intersection-over-union (IoU) is calculated,

$$\mathrm{IoU}(\mathcal{P},\mathcal{G}) = \frac{|\mathcal{P}\cap\mathcal{G}|}{|\mathcal{P}\cup\mathcal{G}|}. \tag{1}$$

If the IoU meets or exceeds a chosen threshold τ, the prediction is counted as a true positive (TP), otherwise it is a FP. Any ground-truth instance without a matching prediction is considered as a false negative (FN).

Average precision at threshold τ is the area under the precision–recall curve:

$$\mathrm{AP}(\tau) = \int_0^1 p_\tau(r)\,\mathrm{d}r, \tag{2}$$

where $p_\tau(r)$ is the interpolated precision at recall (r) with $p_\tau = \mathrm{TP}/(\mathrm{TP}+\mathrm{FP})$ and $r = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN})$.

It is obtained after declaring a detection correct only if its predicted mask overlaps the ground-truth mask by at least τ of their union area. For a single summary metric, AP(τ) is averaged over ten IoU thresholds $0.50 \leqslant \tau \leqslant 0.95$ in steps of 0.05. Thus, a higher AP score typically implies that the model detects more QoIs with more accurate masks, yet absolute values strongly depend on the targeted quantities or metrics. The result is reported separately for bounding-box detection ($\mathrm{AP}_{\mathrm{bbox}}$) and instance segmentation ($\mathrm{AP}_{\mathrm{mask}}$). In addition, the following auxiliary measures are reported:

- AP_{50} and AP_{75}, which are the average precision values at the single IoU thresholds 0.50 and 0.75 respectively. In practical terms, AP_{50} aligns towards more detections, tolerating boundary errors while AP_{75} prioritises precise delineations (for further statistics) of the detected objects. AP_{75} thus considers a detection only when the mask is substantially more accurate, which accounts to a few pixels for medium-sized objects and proportionally less for very small ones.
- Size-conditioned metrics AP_S, AP_M, and AP_L, which restrict evaluation to objects whose pixel areas $A = w \times h$ fall into the COCO size buckets, small : $0 < A(\mathrm{px}^2) < 32^2$, medium: $32^2 \leqslant A(\mathrm{px}^2) < 96^2$ and large: $A \geqslant 96^2\,\mathrm{px}^2$.
- Class-wise AP for the two classes in our benchmark, *piv* (tracer-particle scattering) and *soot* (soot filament structures).

3.5. Models

The 72 models used for benchmark are carefully chosen to study the effect of domain parameters (background variance), QoI object-mix effects, class proportions and fusion methods for objects during synthetic data generation. Training is conducted using Mask R-CNN under identical optimisation settings. Each model corresponds to one synthetic data recipe defined continually to reflect real world situations. First, eight distinct object mix recipes by specifying the percentage shares of experimental soot (ES), CFD soot (CS), generative soot (GS), PIV tracer regions and Out-Of-Plane-Objects (OOPObjs) are designed:

- **Object mix recipes (8)** $\langle ES, CS, GS, PIV, OOPObj \rangle$:
  - Realistic generative: $\langle 0, 0, 25, 50, 25 \rangle$
  - Predominantly generative (least physics): $\langle 0, 0, 50, 50, 0 \rangle$
  - Mixed simulated: $\langle 0, 25, 0, 50, 25 \rangle$
  - Simulation centric: $\langle 0, 50, 0, 50, 0 \rangle$
  - PIV dominant (realistic with object diversity): $\langle 10, 10, 10, 60, 10 \rangle$
  - Fully balanced (best object diversity): $\langle 20, 20, 20, 20, 20 \rangle$
  - Experimental but diversified (most realistic): $\langle 25, 0, 0, 50, 25 \rangle$
  - Purely experimental (least object variance): $\langle 50, 0, 0, 50, 0 \rangle$

Second, for each object mix one of three background families is sampled:

- **Background families (3):**
  - In-scene real PIV windows
  - COCO dataset crops (greyscale)
  - Uniform black canvases

Third, one of three placement policies are applied when compositing objects onto the canvas:

- **Placement policies (3):**
  - `BL` (blend): semi-transparent overlay
  - `RE` (replace): hard overwrite
  - `RA` (random): random combination of the above

This produces a total of 72 recipes (as listed in table 1), all of which have images of dimensions 640×640 in `uint8` format normalised between 0 and 255. During generation, each object is subjected to random augmentations like scaling, rotation and intensity scaling with added synthetic background noise for added robustness.

3.6. Statistics for physical validation

During inference, the model outputs binary masks ($\mathcal{B}(x, y)$) for the predictions. Subsequently, an OpenCV based algorithm extracts the connected contour ($\mathcal{C}$), and calculates the statistics mentioned below and used in section 4.

(i) Area (A): Provided s_x and s_y are the pixel pitch, the zeroth raw moment of the mask gives the area (mm^2),

$$A = s_x s_y m_{00} = s_x s_y \sum_{(x,y)\in\mathcal{B}} 1\,.$$

Table 1. Class shares and assembly options for every synthetic-training recipe (BL = blend, RA = random, RE = replace). Bck. - background, Place. - placement strategy.

ID	ES (%)	CS (%)	GS (%)	PIV (%)	OOPObj (%)	Bck.	Place.
1	0	0	25	50	25	black	BL
2	0	0	25	50	25	COCO	BL
3	0	0	25	50	25	real	BL
4	0	0	50	50	0	black	BL
5	0	0	50	50	0	COCO	BL
6	0	0	50	50	0	real	BL
7	0	25	0	50	25	black	BL
8	0	25	0	50	25	COCO	BL
9	0	25	0	50	25	real	BL
10	0	50	0	50	0	black	BL
11	0	50	0	50	0	COCO	BL
12	0	50	0	50	0	real	BL
13	0	0	25	50	25	black	RA
14	0	0	25	50	25	COCO	RA
15	0	0	25	50	25	real	RA
16	0	0	50	50	0	black	RA
17	0	0	50	50	0	COCO	RA
18	0	0	50	50	0	real	RA
19	0	25	0	50	25	black	RA
20	0	25	0	50	25	COCO	RA
21	0	25	0	50	25	real	RA
22	0	50	0	50	0	black	RA
23	0	50	0	50	0	COCO	RA
24	0	50	0	50	0	real	RA
25	0	0	25	50	25	black	RE
26	0	0	25	50	25	COCO	RE
27	0	0	25	50	25	real	RE
28	0	0	50	50	0	black	RE
29	0	0	50	50	0	COCO	RE
30	0	0	50	50	0	real	RE
31	0	25	0	50	25	black	RE
32	0	25	0	50	25	COCO	RE
33	0	25	0	50	25	real	RE
34	0	50	0	50	0	black	RE
35	0	50	0	50	0	COCO	RE
36	0	50	0	50	0	real	RE
37	10	10	10	60	10	black	BL
38	10	10	10	60	10	COCO	BL
39	10	10	10	60	10	real	BL
40	10	10	10	60	10	black	RA
41	10	10	10	60	10	COCO	RA
42	10	10	10	60	10	real	RA
43	10	10	10	60	10	black	RE
44	10	10	10	60	10	COCO	RE
45	10	10	10	60	10	real	RE
46	20	20	20	20	20	black	BL
47	20	20	20	20	20	COCO	BL
48	20	20	20	20	20	real	BL
49	20	20	20	20	20	black	RA
50	20	20	20	20	20	COCO	RA
51	20	20	20	20	20	real	RA
52	20	20	20	20	20	black	RE
53	20	20	20	20	20	COCO	RE
54	20	20	20	20	20	real	RE
55	25	0	0	50	25	black	BL
56	25	0	0	50	25	COCO	BL
57	25	0	0	50	25	real	BL
58	25	0	0	50	25	black	RA
59	25	0	0	50	25	COCO	RA
60	25	0	0	50	25	real	RA
61	25	0	0	50	25	black	RE
62	25	0	0	50	25	COCO	RE
63	25	0	0	50	25	real	RE
64	50	0	0	50	0	black	BL
65	50	0	0	50	0	COCO	BL
66	50	0	0	50	0	real	BL
67	50	0	0	50	0	black	RA
68	50	0	0	50	0	COCO	RA
69	50	0	0	50	0	real	RA
70	50	0	0	50	0	black	RE
71	50	0	0	50	0	COCO	RE
72	50	0	0	50	0	real	RE

(ii) Aspect ratio (AR): A least-squares ellipse fits the contour; let a_{maj} and a_{min} be its full major and minor axes, then

$$AR = \frac{a_{\mathrm{maj}}}{a_{\mathrm{min}}}.$$

(iii) Solidity (S): With A_{conv} denoting the area of the convex hull ($\mathcal{H}$) obtained using OpenCV,

$$S = \frac{A}{A_{\mathrm{conv}}}, \qquad A_{\mathrm{conv}} = \mathrm{contourArea}\left(\mathcal{H}\right).$$

(iv) Perimeter (l_{contour}): The polygonal arc length of the contour is obtained from OpenCV, giving the perimeter (mm),

$$l_{\mathrm{contour}} = \mathrm{arcLength}\ \left(\mathcal{C}, 1\right).$$

(v) Mean absolute curvature ($\bar{\kappa}$): After adaptive Savitzky–Golay smoothing, the mean absolute curvature (mm^{-1}) is

$$\kappa = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{x_i' y_i'' - y_i' x_i''}{\left((x_i')^2 + (y_i')^2\right)^{3/2}}\right|.$$

(vi) Fractal dimension (D_f): A fast box-counting algorithm counts the number of non-empty boxes $N(b)$ for side lengths $b = 2^k$ (pixels) and four grid offsets. A linear fit over the straightest log–log segment yields

$$\log N(b) = D_f \log(1/b) + c,$$

so that D_f is the negative slope. A line has $D_f = 1$ and a filled circle has $D_f = 2$. Smooth contours thus give a low D_f in the limit approaching unity and highly corrugated contours approach the limit of two.

4. Results and discussion

4.1. Model benchmarking and validation

4.1.1. Global model performance

For each of the 72 training runs, the checkpoint with the highest class-specific precision, AP_{soot}, is obtained thereby benchmarking every model at its best soot-segmentation iteration. Figure 4(a) summarises three IoU-stratified metrics: AP_{50}, overall AP, and AP_{75}. Across almost all models AP_{50} dominates, approximately twice the overall AP and three times AP_{75}, revealing that mask quality degrades sharply under stricter overlap thresholds and that metric rankings diverge (e.g. models 3 and 9 lead on AP_{50} at 37%, whilst six other recipes share the best AP_{75} at 7%; only model 3 tops the overall AP at 14%). Figure 4(b) decomposes the same runs by object size. Large structures (soot filaments and PIV) account for roughly 60% of the total AP, whereas small soot filaments seldom reach 5% (AP_S), producing a tenfold gap (e.g. models 3, 9, 33 and 57 achieve $AP_L > 35\%$ but $AP_S < 4\%$); medium-size performance spans 4%–18%. Together, the two figures show that (i) global AP is dominated by easy to segment, large objects, (ii) IoU-stratified metrics are only weakly correlated, and (iii) model rankings depend strongly on both threshold and size regime. Moreover, although bounding-box performance somewhat correlates with soot-mask AP (Pearson $\rho = 0.71$), it overstates by an average inflation gap of ~ 10 pp. This bias widens for models with higher AP_{bbox} score. This divergence is visualised in supplementary methods and materials figure S2, confirming that AP_{bbox} is an unreliable proxy for pixel-level soot quality and that class-specific segmentation metrics remain essential. These findings argue for a multi-term scoring function and a more robust model-selection framework when training on synthetic data [52], so that fine soot structures are not obscured by coarse but numerically dominant detections.

To assess the underlying cause for depletion and sensitivity of the global AP metric score, the composite figure 5 showcases the effect of specific FNs and FPs. Focus is given to the background and image fusion strategies, due to its high sensitivity towards false detections. For the model applicability, it is highly beneficial to evaluate real-world effects of the model's detection capability, rather than relying on AP scores that reduces each detection event to a percentage value [37]. Figure 5(a) indicates the trade-off surface for FNs and FPs for all models, with minimising both scores offering higher quality and more robust predictions. The COCO background, with relatively weak performance, occupies the top right quadrant. Conversely, the models trained on real and black background show improved performance with low FN and FP values. The lowest error rates are demonstrated in models trained on real background image patches, irrespective of the image fusion strategy (i.e. data points close-to origin). A considerable number of models are prone to fewer FN rates, but more hallucinations as seen from the top left quadrant. Figure 5(b) indicates the mean error for FN and FP by background. COCO background produces again approximately double the error rate compared to real and black backgrounds, in which FN contributes to approximately 60% of the total error rate. This suggests that a higher synthetic domain parameter can overwhelm the model, making it prone to miss instances and requiring more domain adaptation. The swarm-and-box plot (figure 5(c)) indicate the error rate distribution for all models. As expected, COCO backgrounds realise the weakest performance and yield a higher variability (0.12 median for FN versus 0.08 for FP) signalling lower robustness. For FN rates, both real and black background have narrower spread and improved robustness. The opposite is seen for FP rates, where real and black backgrounds contribute to higher spread. These uncertainties and variances in performance metrics render corresponding optimisation processes non-trivial. The large number of hyper-parameters and synthetic training data generation options inherently call for swarm-based approaches or Bayesian uncertainty methods. The current systematic variation to identify the best strategy is rather costly, yet clear trends are evident and strategies to compile synthetic training data can be derived.

4.1.2. Composite soot-segmentation score

Considering the need for a single figure-of-merit that (i) focuses on soot mask quality, (ii) rewards high-precision boxes in combination with more accurate masks, and (iii) penalises both types of classification

Figure 4. Benchmark performance of 72 trained segmentation models. (a) IoU stratified average precision: each stacked bar reports the headline AP, AP$_{50}$ and AP$_{75}$. (b) Size stratified average precision: the same models decomposed into large (AP$_L$), medium (AP$_M$) and small (AP$_S$) objects.

error, a weighted soot-centric composite score is defined as

$$\mathcal{S}_{\text{soot}} = 0.50\,\widehat{\text{AP}}_{\text{soot}} + 0.125\,\widehat{\text{AP}}_{75} + 0.125\,\widehat{\text{AP}}_{\text{size}} + 0.125\,(1 - \text{FN}) + 0.125\,(1 - \text{FP})\,, \tag{3}$$

where $\widehat{\text{AP}}_{\bullet} = \text{AP}_{\bullet}/100$ and $\widehat{\text{AP}}_{\text{size}} = 0.50\,\widehat{\text{AP}}_{\text{L}} + 0.30\,\widehat{\text{AP}}_{\text{M}} + 0.20\,\widehat{\text{AP}}_{\text{S}}$. The highest weight is given to $\widehat{\text{AP}}_{\text{soot}}$ as the primary objective is to segment soot structures with high quality masks. These include the strict general metric, $\widehat{\text{AP}}_{75}$, which is given the second-highest weight as it determines the overall mask quality. In addition, the size-weighted metrics ($\widehat{\text{AP}}_{\text{size}}$) are considered to include the size bias of small, medium, and large object scores. Their individual weights are chosen to represent the trends in figure 4. The subsequent terms, FP and FN are included to inject an explicit penalty for the two critical failure modes that a mean-IoU metric alone can hide. The normalisation to $[0, 1]$ puts all added terms on a

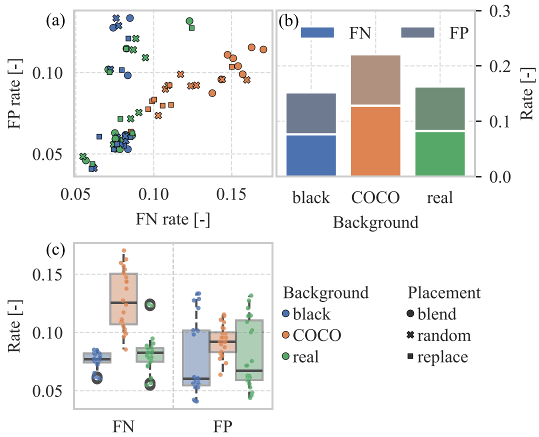

Figure 5. Error-trade-off dashboard (false negatives vs false positives). Per-model FN-FP scatter coded by background type (colour) and placement policy (marker) is shown on subplot (a). Mean FN and FP rates for each background is shown on subplot (b). Swarm/box plots of FN and FP distributions is shown on subplot (c). Models trained with COCO backgrounds yield the highest error rates, whereas real and black backgrounds produce lower, more balanced uncertainties.

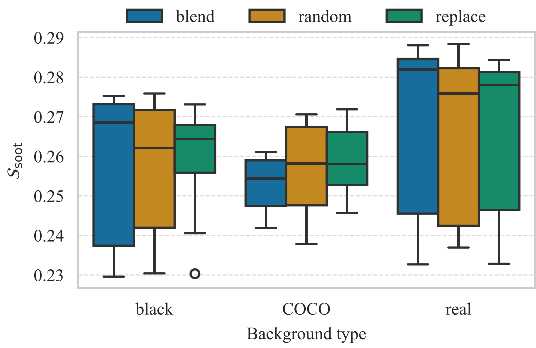

Figure 6. Composite soot-segmentation score by background and placement strategy. Box-plots show that real-image backgrounds consistently outperform black and COCO, while blend placement gives the highest median scores within each background family.

common scale and the overall score remains dimensionless and interpretable as a weighted mean percentage. For the remaining plots this score is used.

Figure 6 shows the composite score dispersion by background and placement. Background contributes to most of the variance, raising the median score from ∼0.25 to ∼0.26 and to ∼0.28 for COCO, black and real backgrounds, respectively. It further shows that irrespective of image fusing strategy, choosing real background adds ∼0.02 to the composite score. In general, fusion strategy effects remain consistent with *blend* performing the best and *replace* performing the worst (except for COCO). It is also worth noting, that background combination of different categories (i.e. a mix of *black, COCO*, and *real*) does not improve the detection quality for the current data. Thus, real background with objects partially or completely blended into it performs best. Although the plot only has a dynamic range of 0.06 for the composite score, it facilitates resolving such subtle, yet significant differences between data generation factors.

Figure 7 summarises how the composite score responds to the proportion of each synthetic QoI object class (i.e. ES and OOPObjs, computational and generative). It is worth noting that the PIV tracer particle class is included in all the datasets with a minimum share of 20%. The heat-maps illustrate that adding 15%–25% synthetic soot (CS or GS) to 10%–25% ES yields the highest scores. It further shows that CS and GS prove interchangeable when kept within the 20%–30% contribution band. This highlights that object diversity is key and that maintaining the share of ES between 10% and 25% provides enough texture for the model to steer towards better validation performance on real experimental data. This offers distinct benefits in terms of autonomous data labelling and allows reductions in manual overhead as well as optimisation towards domain-invariance. Simultaneously, unbalancing the QoI object class distribution, e.g. to individual shares of 50%, generally lowers performance. This indicates overfitting to artefacts of the synthetic training data. The interaction of the QoI object class balance with

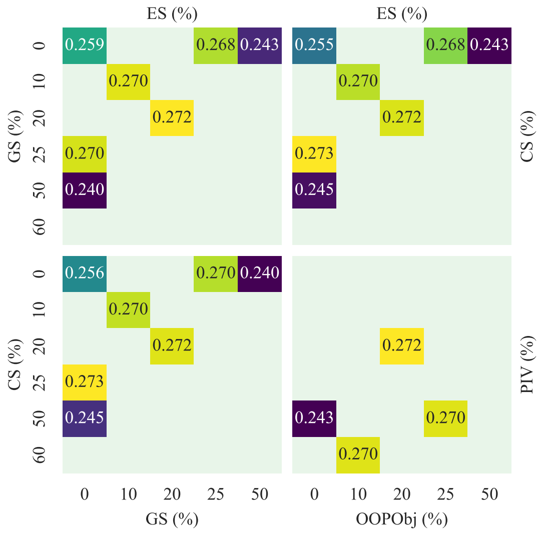

Figure 7. Ablation of object proportions. The heat-map indicates the influence of each QoI on model performance quantified through the composite score. Diversified QoI portfolios promote model generalisation and thereby its segmentation performance.

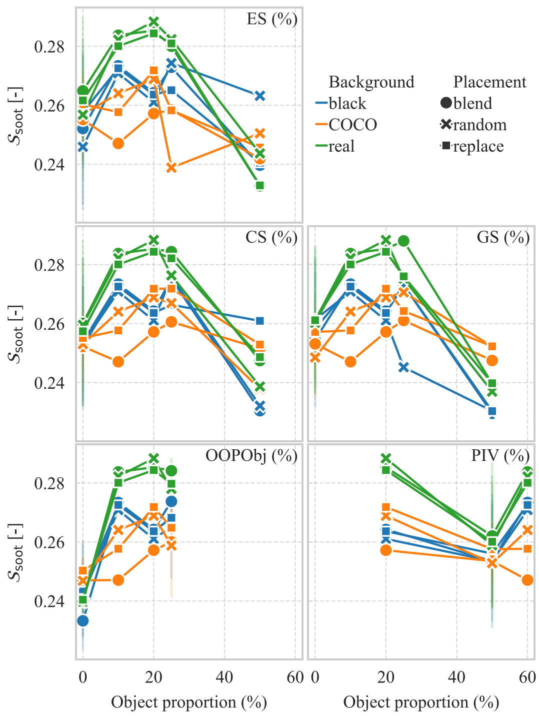

Figure 8. Marginal impact of each class share on composite score (lines = mean, band = ±1 SD). The *x*-axis denotes the percentage proportion of each QoI.

the background and image fusion strategy is depicted in figure 8. The partial-dependence curves confirm these trends for each class individually. Every curve exhibits a peak around 20%, flanked by steeper declines on either side. Real-image backgrounds lift the score by ∼0.02, while the choice of fusion strategy has negligible effect at the optimum. The data further show that COCO backgrounds exhibit minimal sensitivity to class proportions due to the higher semantic variance. However, this robustness comes at the cost of degraded soot-segmentation accuracy. The shaded ±1 standard deviation envelopes in figure 8 widen markedly when a class is absent (i.e. 0% share). This signals heightened volatility, and, consequently, costlier hyper-parameter search in those regimes. The main take away is the necessity for distinct object diversity when dealing with domain randomised datasets.

Table 2. Weight vector(s) $\mathbf{w} = \langle w_{\text{soot}}, w_{75}, w_{\text{size}}, w_{\text{FN}}, w_{\text{FP}} \rangle$ used in figures 6–9 (coefficients sum to 1).

Set	w_{soot}	w_{75}	w_{size}	$(1-\text{FN})$	$(1-\text{FP})$
S1 (baseline)	0.50	0.125	0.125	0.125	0.125
S2 (sharp-contour)	0.45	0.25	0.125	0.10	0.075
S3 (size-balanced)	0.45	0.05	0.30	0.10	0.10
S4 (safety-bias)	0.45	0.10	0.10	0.25	0.10
S5 (precision-bias)	0.45	0.10	0.10	0.10	0.25

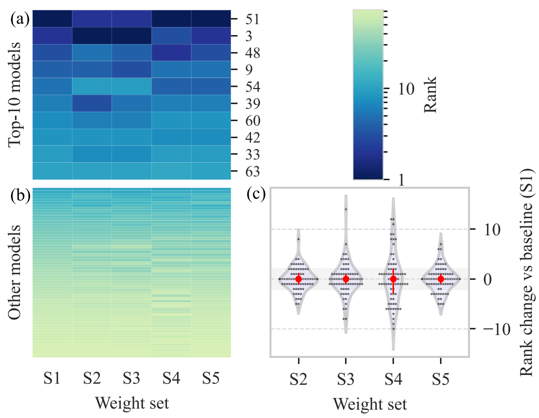

Figure 9. Model ranking stability under weight perturbations. Subplot (a): Top 10 models maintain their ranks even with significant weight shifts showcasing model robustness. Subplots (b) validates the same for other models. In subplot (c), the weight-sets (S2–S5) are compared to the baseline S1 demonstrating that the ordering is essentially insensitive to large weight rebalancing.

Table 3. Top Mask R-CNN configurations. Percentages denote the share of each object class in the synthetic-training set. Placement: BL = *blend*, RA = *random*, RE = *replace*. All these models are using real backgrounds from the experimental data.

Rank	ID	ES (%)	CS (%)	GS (%)	PIV (%)	OOPObj (%)	Plmt.	Bck.
1	51	20	20	20	20	20	RA	real
2	3	0	0	25	50	25	BL	real
3	48	20	20	20	20	20	BL	real
4	9	0	25	0	50	25	BL	real
5	54	20	20	20	20	20	RE	real

4.2. Sensitivity analysis

To verify that the conclusions do not depend on a particular choice of coefficients, the composite score is recomputed under four alternative weight vectors, each stressing one auxiliary component while keeping the normalisation to unity and baseline weight of AP_{soot} close to 0.5 as shown in table 2. The corresponding effect of these weight perturbations on the rank of the models based on composite score is depicted in the figure 9. The first heat-map (figure 9(a)) indicates the rank variations of top-10 models under the five weight sets. The rank matrix for other models is indicated in the corresponding compressed plot (figure 9(b)). Even with substantial changes in weight distribution, only slight changes in the model ranks are evident. The best model (ID 51) never drops below rank 2 and the 10th best never rises above rank 9. The other models in the compressed plot also remain reasonably stable. In the violin plot (figure 9(c)), where each weight-set (S2–S5) is compared to the baseline ranking S1, this is again verified as the median Δr sits exactly on the dashed zero line, confirming that the typical model remains robust to particular requirements of the user. The red IQR bar for all sets spans at most one tick above and below zero; therefore 50% of the models migrate by no more than one place. Moreover, 95% of the models are almost entirely confined to ± 5 ranks for S2, S3 and S5, and to ± 8 ranks for the safety-biased set S4. All these observations underscore the robustness of the composite score and the generalisation of the model towards significant changes towards usability.

Based on the thorough evaluation, the top 5 models are listed in table 3. It is clear, that the models with the most object diversity perform best and are generally more invariant towards the fusion methodology. Although recipe ID 51 tops the ranking with a random placement policy, the composite score margin is low and three of the top five models use a blending strategy and thus following the trend

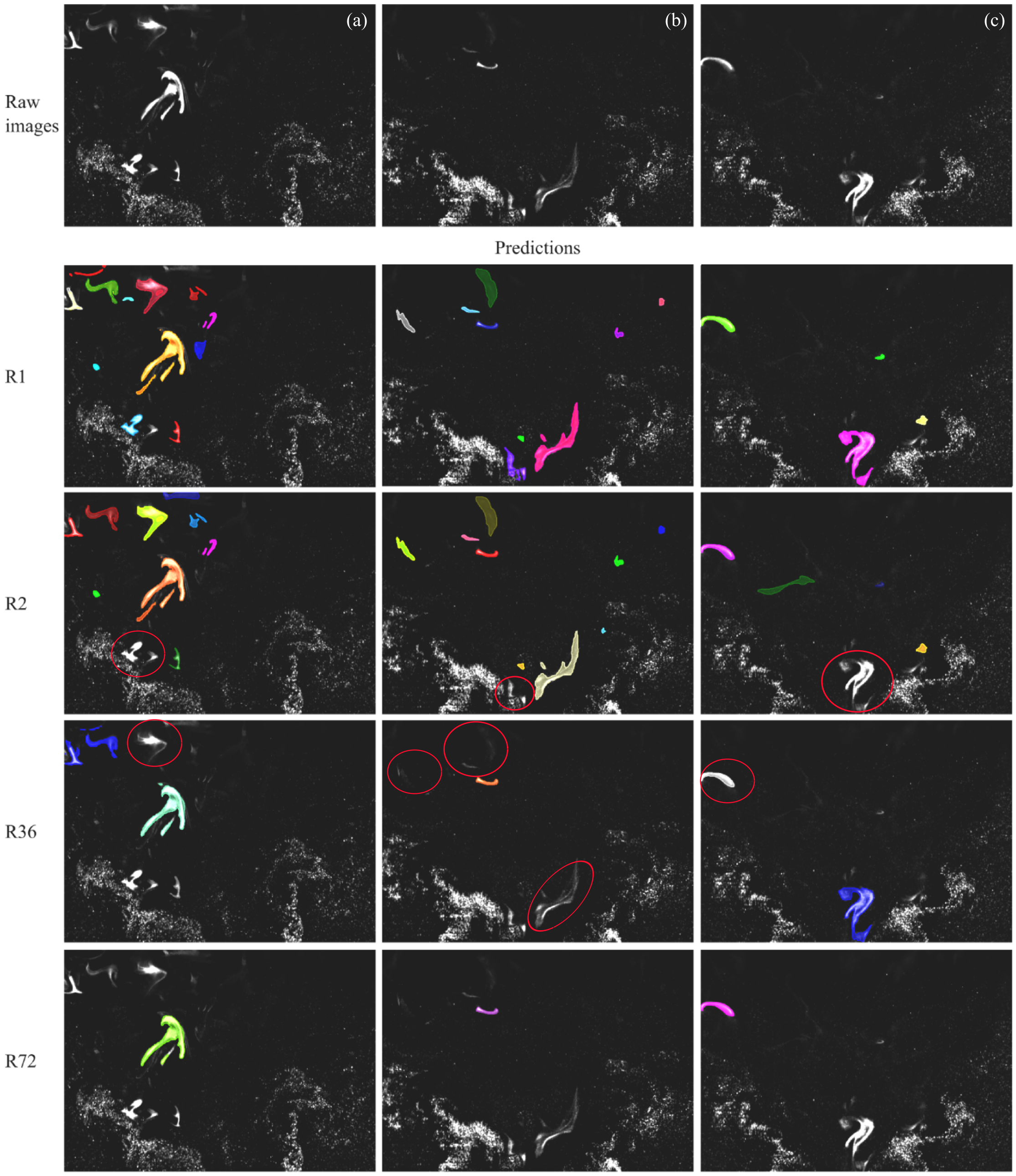

Figure 10. Instantaneous, brightness adjusted raw images (top row) and corresponding instance segmentation predictions (coloured masks) from models of ranks 1, 2, 36 and 72. The models predict reasonable masks appropriate for further statistical analysis. Subtle differences (marked in circles) can be seen even between the adjacent ranked models (ranks 1 and 2), while significant degradation of performance is evident when the rank goes further down.

shown in figure 6. Photo-realistic backgrounds from the experiments are used by all the top 5 models, indicating that black backgrounds fail to provide enough semantic information, while highly random COCO backgrounds can overwhelm the model or ease over-fitting to the synthetic training data. A visual interpretation of these metrics as predictions on real images is shown in figure 10. Subtle, yet statistically significant differences can be seen in the predictions of the top two models even though their composite score is similar. As the rank goes down, these differences become prominent as missing and incomplete predicted masks, underlined by the composite score.

4.3. Physical validation

The best model (ID 51) is used for inference on the real images from the experiments (∼24k). The model produces segmentation masks for each detected soot instance, as shown in figure 10. The predictions are used for further conditional statistics. At this point it is worth noting, that the obtained

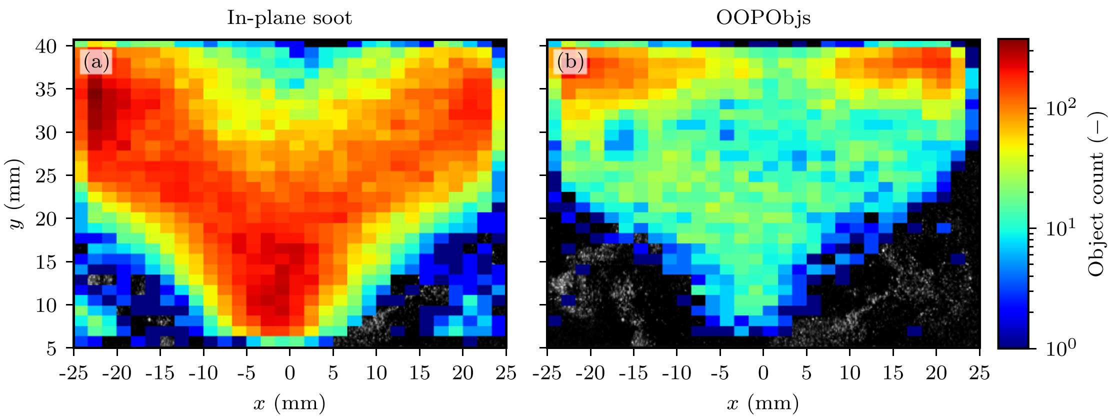

Figure 11. Distribution of the number of detections (in-plane soot and OOPObjs) within the combustor geometry shown as a heat-map.

instance segmentation masks (in uint8) are used to condition the raw images restoring the original dynamic range. The statistical analysis pipeline is defined as:

- Masks, confidence scores, and bounding box coordinates, generated by the model, are stored.
- Duplicates and bad predictions are filtered based on a predefined confidence (0.5 out of 1.0)
- The binary mask and bounding box predictions are used to extract the objects from the real image. These objects, along with other mentioned information, facilitate to relocate the object outside of the inference pipeline from the data container. This maintains the original dynamic range and spatial resolution.
- For each detected and extracted object, quantities like mask surface and contour statistics are calculated and stored along with the metadata of the source image.
- Based on a K-means [62] clustering algorithm with 9 clusters, the detected objects are classified into in-plane soot and OOPObjs based on their mean intensity variance for further accurate and contextual conditional statistics.

A total of ∼85k objects are detected by the model of which ∼69k being classified as in-plane soot and ∼16k as OOPObjs. The distribution and density of these detections are illustrated in figure 11. Each image is divided into 30 × 30 grid cells to conduct a spatially resolved statistical analysis. The FPs (PIV classified as soot) contribute <0.5% of the total in-plane soot objects which is well below accepted noise thresholds. Roughly 50% of that (∼150 instances) is attributed to soot masks connected to high intensity PIV regions, that are omitted in the statistical analysis. Some artefacts such as filled masks of hollow or ring-like soot structures and mask overlaps are observed, yet remain statistically insignificant due to their low occurrence and active filtering (< 1%). This deficiency can be attributed to the used 'polygon' mask format in the model training. An illustration of the different zones mentioned in the statistical analysis below is depicted in figure 12. As seen in figure 11(a), the in-plane soot clusters along the outer edge of the IRZ. By contrast, hot spots of OOPObjs are found in the top region of the interrogation region, at radial positions away from the symmetry axis (figure 11(b)) at $|x| \geqslant 10$ mm. This is, to some degree, expected due to the high out-of-plane flow motion in this region. Figure 13 illustrates soot occurrence frequency (frame intervals) conditioned into super-pixels, localised using soot filament centroid coordinates. In each super-pixel cell, the corresponding frame IDs of the detected soot are evaluated to find the intervals (i) of soot absence, where the coefficient of variation ($\mathrm{CV} = \sigma/\bar{i}$) quantifies its intermittency. Subsequently, the mean ($\bar{i}$), median ($\tilde{i}$) and standard deviation (σ) are calculated (figures 13(a)–(d)). Large values (CV~ 1.7) are observed along the IRZ indicating high intermittency (short, intense bursts of soot appearance that are separated by long soot free intervals), corresponding to the long residence times mentioned in [15]. Lower mean ($\bar{i} \sim 150$) and median ($\tilde{i} \sim 50$) with higher standard deviation ($\sigma \sim 400$) confirms that soot detections (emissions) manifest as short, intense bursts separated by extended quiescent periods as also reported by Stöhr *et al* [15].

Figures 14 and 15 illustrate spatial distribution heat-maps of the geometric as well as interface properties such as area (A), aspect ratio (AR), solidity (S), perimeter (l_{contour}), mean absolute curvature (κ) and fractal dimension (D_{f}) of the detected in-plane soot structures. Each image is divided into 30 × 30 grid cells, clustering 20 × 14 pixels^2, where within each cell statistical moments such as mean ($\bar{x}$),

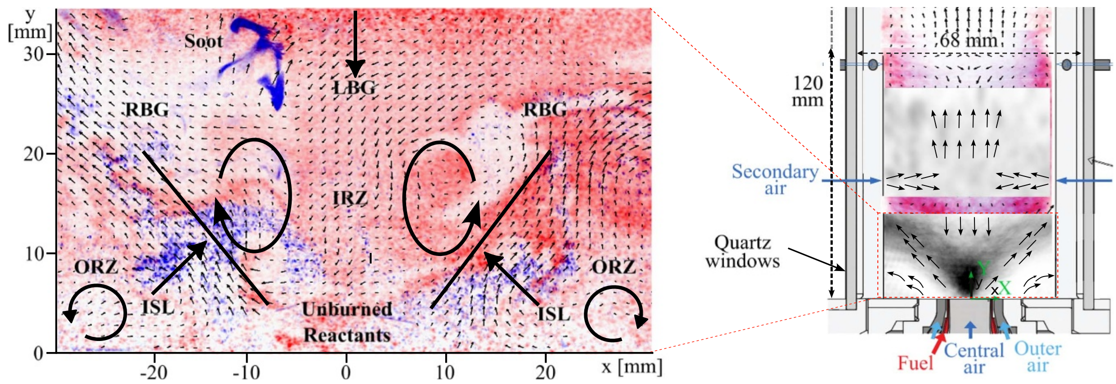

Figure 12. Distribution of different zones in the combustor (derived from Stöhr *et al* [15]). The Mie scattering field of PIV tracer particles and soot filaments is shown in blue, superimposed onto the OH-PLIF scalar field in red. Rich burned gas (RBG) regions are marked by low OH signal levels. The lean burned gas (LBG) regions exhibit medium levels of OH. The inner shear layer (ISL) separates the inner recirculation zone (IRZ) from the unburnt reactants marked by high seeding density in blue. The instantaneous ISL is inherently highly convoluted and transient. Its statistical location is illustrated by the two lines. The outer recirculation zone (ORZ) is at radial positions with $|x| \geqslant 20$ mm away from the burner centre.

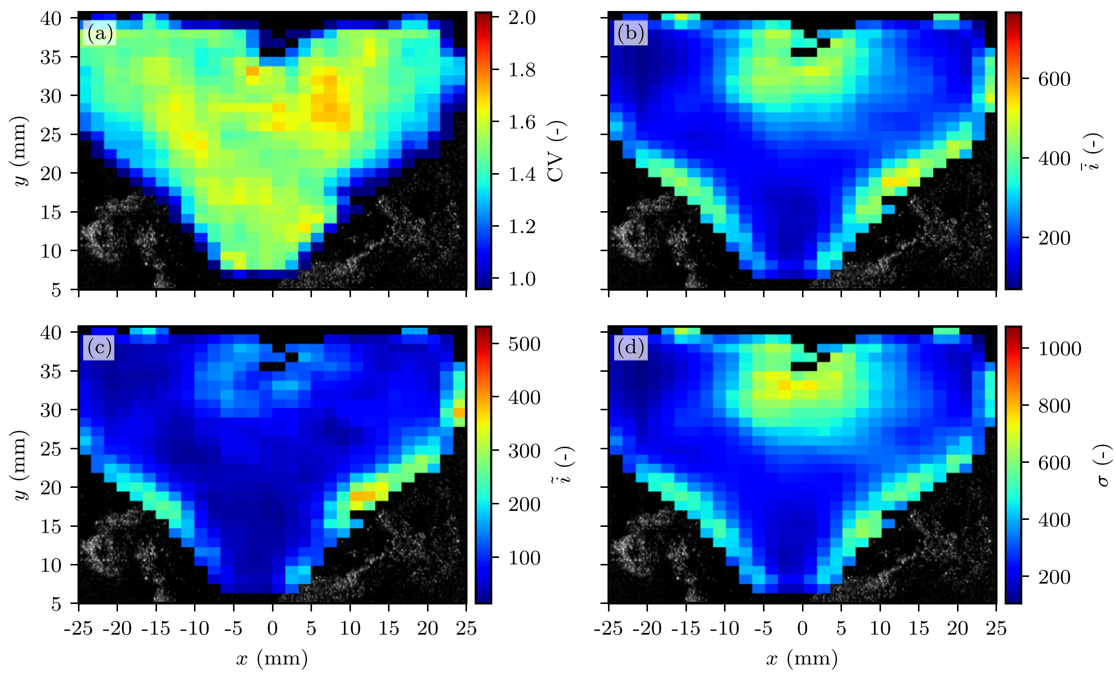

Figure 13. Illustration of soot occurrence frequency conditioned into super pixels. Coefficient of variation (CV) shows higher intermittency in the IRZ corresponding to the long residence times. Lower mean ($\bar{i} \sim 150$) and median ($\tilde{i} \sim 50$) with higher standard deviation ($\sigma \sim 400$), with high CV (~ 1.7) confirms intermittent soot detections (emissions).

median ($\tilde{x}$), standard deviation (σ), kurtosis (β_2) and skewness (γ) are evaluated. The accumulative area moment visualisation allows to infer the underlying positively skewed distribution. The mean area (figure 14(a)) peaks ($A \simeq 5\,\text{mm}^2$) in the region $10 \leqslant |x|\,(\text{mm}) \leqslant 15$ and $15 \leqslant y\,(\text{mm}) \leqslant 30$, i.e. transport into the IRZ, where the long residence time of rich burned gas (RBG) pockets facilitates soot formation and growth [15, 17]. In the IRZ—inner shear layer (ISL) contact points close to the burner nozzle, i.e. $5 \leqslant |x|\,(\text{mm}) \leqslant 10$ and $5 \leqslant y\,(\text{mm}) \leqslant 15$, the standard deviation of the soot filament area (see figure 14(c)) shows hot spots. This can be attributed to the transport of large soot filaments through the IRZ to the contact points, where the high shear in the ISL strips and fragments the filaments. By contrast, mean and median values for aspect ratio (see figure 14(f)) peaks along the ISL between the reactants and the transported rich burnt gases containing soot filaments. The higher strain rates stretch the newly generated soot filaments azimuthally, thereby increasing their aspect ratio (AR) up to 3.6 in the mean. This also corresponds with the area statistics, where the positive skewness along the IRZ (see figure 14(e)) indicates compact soot dominance, while the lower values along the ISL indicate larger and stretched soot filaments. The aspect ratio standard deviation (figure 14(h)) co-peaks with

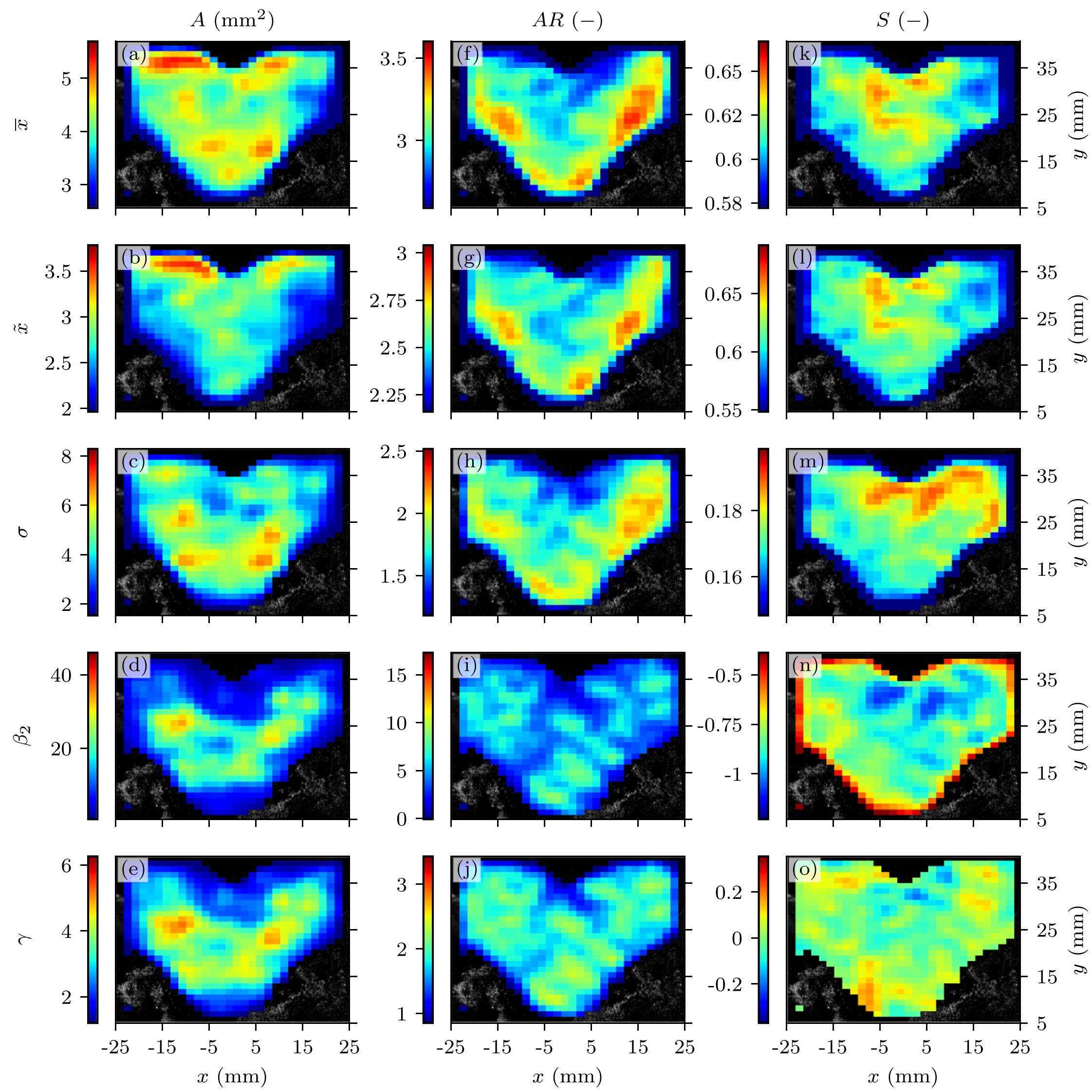

Figure 14. Spatial distribution heat-map showing how geometric properties such as area (A), aspect ratio (AR) and solidity (S) of the detected objects vary across the field of view. The image is divided into a 30 × 30 grid cells. Within each cell, five statistical moments are computed and shown by colour: mean ($\bar{x}$), median ($\tilde{x}$), standard deviation (σ), kurtosis (β_2) and skewness (γ).

mean values, revealing that the shear layer not only elongates filaments on average, yet also introduces large shape variability. The central region remains predominantly empty (low values of aspect ratio) of stretched soot structures indicating that the convex soot filament hulls are preferentially compacted within the IRZ. Solidity, somewhat inherently, inversely correlates with the aspect ratio and area as it peaks ($S \simeq 0.65$) in the IRZ where more compact and roundish soot structures are evident. The solidity reduces towards the IRZ edge (see figure 14(k)) and near the ISL. This is again attributed to the stretching of soot surfaces by high shear. The higher standard deviations (figure 14(m)) and kurtosis (figure 14(n)) indicate that the soot formation and erosion depends highly on the intermittent flow variations in LBG towards the IRZ. Moreover, the accumulative mode visualisation shows that the aspect ratio and, in particular, the solidity are distinctly closer to normal distributions in comparison to the area statistics.

As seen in figure 15, the mean perimeter peaks ($l_{\text{contour}} \simeq 16$ mm) correlate with the area and, inversely, with the solidity statistics as described above. The highest l_{contour} variance is evident in the valley of the IRZ close to the flame root, i.e. $-10 \leqslant x\,(\text{mm}) \leqslant 10$ and $10 \leqslant y\,(\text{mm}) \leqslant 25$ (see figure 15(c)). This can be attributed to the intense stretching and fragmentation of soot filaments when in contact with the high shear of the ISL. As seen from figures 15(d) and (e), the higher values for kurtosis ($\beta_2 \simeq 10$–15) and skewness ($\gamma \simeq 2.5$–4) along the ISL validates the geometrical changes in this region with more compact soot in the IRZ. The high variance of l_{contour} above the flame root aligns well with the region of highest absolute soot flux, as analysed by Stöhr et al [6]. The intense shape variations can,

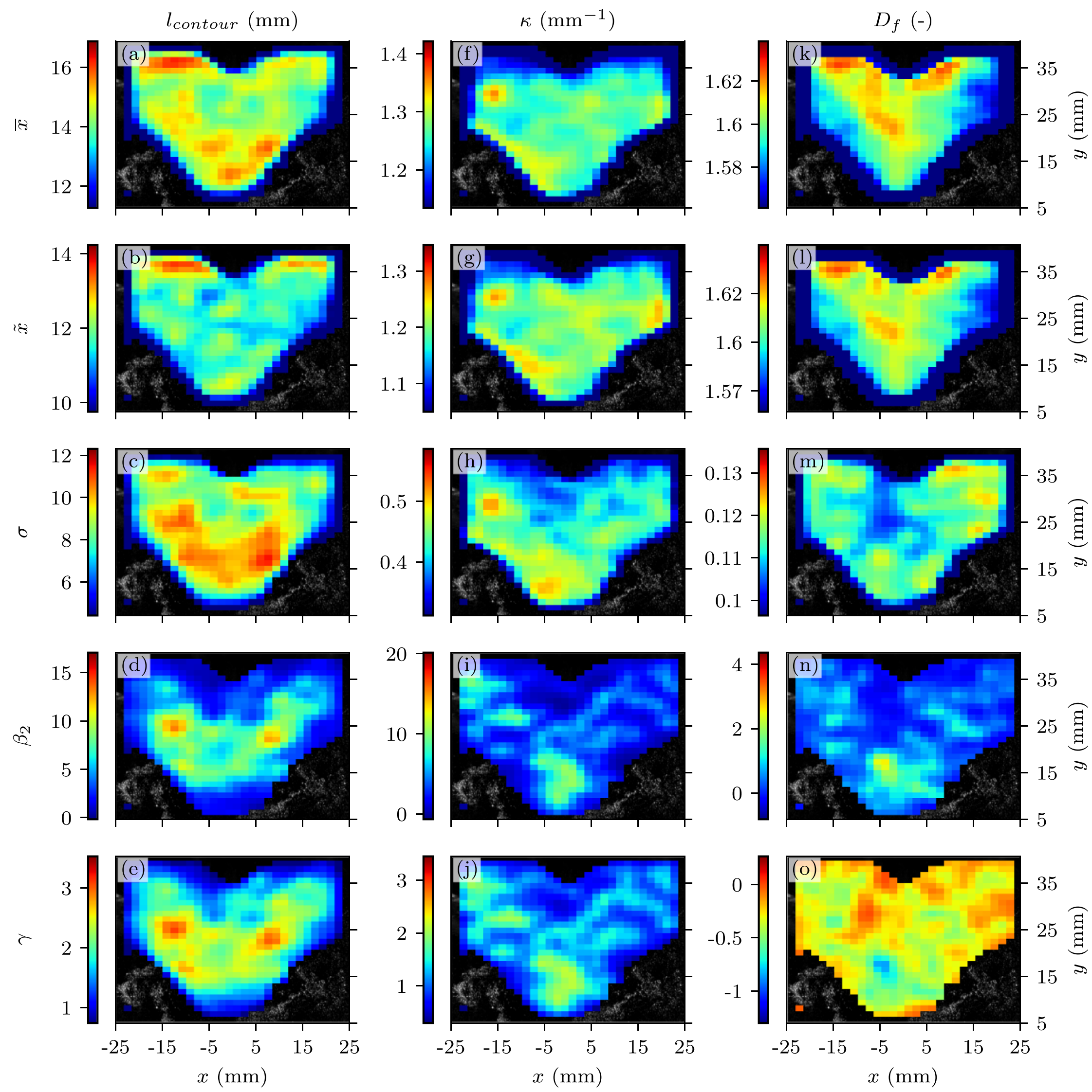

Figure 15. Spatial distribution heat-map showing how interface-level properties such as perimeter (l_{contour}), mean absolute curvature (κ) and fractal dimension (D_{f}) of the detected objects vary across the field of view. The image is divided into a 30 × 30 grid cells. Within each cell, five statistical moments are computed and shown by colour: mean ($\bar{x}$), median ($\tilde{x}$), standard deviation (σ), kurtosis (β_2) and skewness (γ).

arguably, be attributed to high soot filament growth / agglomeration and oxidation rates. The mean absolute curvature peaks at $\kappa \simeq 1.35\,\mathrm{mm}^{-1}$ in the outer edge of the ISL (see figure 15(f)) and is lowest in the inflow region of the IRZ that is accompanied with low strain. The variance in the soot filament bounding contour curvature (figure 15(h)) is highest in the inner vortex core, i.e. $-10 \leqslant x\,(\mathrm{mm}) \leqslant 10$ and $5 \leqslant y\,(\mathrm{mm}) \leqslant 15$, where the soot filaments experience the strongest acceleration and thus folding and stretching. This is further illustrated by the mean fractal dimension (see figure 15(k)), that decreases gradually from ∼1.63 to ∼1.55 towards the ISL, evidencing high stretching due to shear. The standard deviation of D_{f} (figure 15(m)) is highest in the region of high curvature and contour length fluctuations as well as soot out- and inflow region ($12 \leqslant |x|\,(\mathrm{mm}) \leqslant 25$ and $20 \leqslant y\,(\mathrm{mm}) \leqslant 35$). The spatially resolved statistics of κ and D_{f} illustrate the effectiveness of the turbulent flow field to fold and fragment soot filaments. Pronounced shape fluctuations are particularly distinct in regions where the flow field exhibits generally high vorticity and strain. In the IRZ with moderate flow field gradients, the soot filaments reside in a relatively low strain region, preferential for soot formation with positive soot flux [6], favouring the growth to relatively large and smooth structures and high D_{f}. The soot filaments are then transported towards the flame root and ISL, where the high shear yields elongated and fragmented soot filaments with high $\overline{\kappa}$, κ', and AR accompanied with relatively low D_{f}. This region was previously identified as preferential for soot oxidation with negative soot flux [6]. It is further worth noting, that the regions with highest curvature and D_{f} kurtosis (i.e. figures 15(i) and (n)) spatially overlap with the most negative soot fluxes, i.e. oxidation. Overall, these metrics illustrate the effects of turbulent transport in the soot formation and oxidation region, where folded and fragmented soot structures with low curvature are intermittently formed in the IRZ. The provided insights indicate that capturing related turbulence–chemistry-phase interactions accurately is essential to numerically predict soot distributions, especially including an adequate description of their oxidation, subsequently leading to particulate emissions. Experimentally empty soot regions (conical jet region and combustor corners) remain relatively empty in the statistics indicating that the ∼0.5% FPs do not bias the statistics. The consistency and agreement of the metrics with the previous studies indicate general as well as physical validity of the DL-detection model and its adaptability towards real data. Moreover, although the development of such DL-detection models can be costly, the improved general applicability and feasibility of domain-invariance can accelerate downstream data analysis processes [37, 63]. For example, the same statistical analysis is also conducted on the OOPObjs and provided in the supplementary materials and methods. OOPObjs are predominantly detected in the top part of the interrogation regions where they are rather large, exhibit a low aspect ratio and curvature, and a fractal dimension approaching $D_\mathrm{f} = 1.7$. Thus, OOPObjs can be categorised as kind of diffuse structures, while in-plane soot filaments exhibit fine-scale features.

5. Conclusions

The overarching objective of the current work is to develop DL-based segmentation strategies to support the analysis of image-based diagnostics and ease access to high-level statistical data. For this purpose, the present study evaluates strategies for training on purely synthetic data that are created on the basis of domain randomisation. In this context, end-to-end feasibility of a purely synthetic and automatic pipeline with minimal annotation effort, yet high-accuracy masks and reproducible workflow has been demonstrated by segmenting soot filaments in challenging PIV images. A formulated physics-aware composite score, based on soot-specific average precision score, strict IoU, size and FN/FP penalties effectively rank the best models for soot filament detection. This allows to avoid the bias introduced by single metric model performance evaluation and addresses the challenges associated with the model transfer from synthetic to real data (e.g. low $\mathrm{AP}_\mathrm{soot}$ metrics). The robustness of the developed score is validated by means of a sensitivity analysis, where the top-10 model rankings held their position well. Domain parameters like background tend to significantly affect the model performance, where the use of real background patches in the synthetic training data generation process facilitates efficient model learning, striking a promising accuracy-robustness balance with reduced over-fitting risks. On the other hand, black background scenes regularise and natural scenes from the COCO dataset enforce generality. Blending the QoI objects, here soot structures, with the background preserves the realistic luminance and gradients enhancing the segmentation accuracy. Predominance of a single QoI object class in the dataset hinders model learning while diversity (15%–25% per database class) promotes generalisation. This allows to supplement the annotated dataset of QoI objects with data from numerical and generative sources that are comparatively trivial to segment due to the absence of measurement noise. In general it can be concluded that the initial segmentation accuracy somewhat suffers from extensive usage of artificial objects in the synthetic dataset, yet it eases model fine-tuning, e.g. through active feedback loops, for further measurement, while minimising manual annotation effort. Moreover, a minimal usage of artificial object sources enhances the model performance when transitioning from synthetic to real data. Residual limitations (FPs and mask artefacts) could possibly be mitigated using alternative network architectures (e.g. uncertainty-aware learning) and advanced training data-generation strategies which are subject to future work. The best model is physically validated by downstream advanced conditional statistics on the DL-segmented soot filament structures and comparison with previous studies. The fidelity of the obtained soot filament masks offers access to high-level statistical properties that shed light on the complex interactions between the turbulent flow field and soot formation, transport and oxidation regions. The observed structural differences between in-plane soot filaments and OOPObjs can be used to further fine-tune the detection models by additional class separation. The latter inherently comes at the cost of a pre-defined separation of in- and out-of-plane soot filaments. At this point it is worth noting that physical corrections implemented in the experimental procedures (e.g. light sheet homogeneity / pulse fluctuations) or signal separations (e.g. spectral / temporal filtering) should always be preferred. However, if such methods can experimentally not be implemented or are proven insufficient, the present work demonstrates the effectiveness of a comprehensive, automated, parametrised and reproducible training pipeline coupled with a physics-aware score for building trustworthy segmentation models, here of soot filaments in Mie scattering images for PIV. It further builds the foundation for model optimisation towards i) domain-specific detection accuracy or ii) domain-invariance by means of transfer learning on the basis of an expanded database containing data from various experimental and numerical sources. It is worth highlighting that the developed workflow is transferable to other image-based diagnostics as readily demonstrated for droplet [21] and spray segmentation [37]. By employing such data analysis tools, the access to high-level statistics is eased, thereby facilitating quantitative descriptions of turbulence-scalar and scalar-scalar interactions and enhancing validation of respective numerical models.

Data availability statement

All data, code, and trained models supporting this study are openly available via the DaRUS repository (University of Stuttgart). To accommodate different user entry points, from full pipeline reproduction to targeted use of specific data types, we provide a complete project repository as well as separate, thematically structured datasets and the best-trained model under the following DOIs:
Full repository: Code, Data and Models: https://doi.org/10.18419/DARUS-5512.
Trained model: https://doi.org/10.18419/DARUS-5184.
Database for soot filaments: https://doi.org/10.18419/DARUS-5180.
Database for tracer particles: https://doi.org/10.18419/DARUS-5181.
Database for OOPObjs: https://doi.org/10.18419/DARUS-5183.
Database for backgrounds: https://doi.org/10.18419/DARUS-5182.

Acknowledgments

In fond memory of our esteemed colleague and friend, Michael Stöhr, whose invaluable scientific exchange and provision of validation data were instrumental to this work. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number: 456687251. Part of the used experimental and LES data has been derived from Fuel Injector Research for Sustainable Transport (FIRST) under Contract No. 265848 and SOPRANO under Grant No. 690724. The computing facilities have been provided by SimTech, funded by the DFG under Germany's Excellence Strategy—EXC 2075 - 390740016. Special thanks also to Peter Gerlinger and Martin Grader, DLR Stuttgart, for providing the numerical soot structures images.

Funding

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project: 45 668 7251. DFG under Germany's Excellence Strategy—EXC 2075 - 39 074 0016. Part of the used experimental and LES data has been derived from Fuel Injector Research for Sustainable Transport (FIRST) under Contract No. 265 848 and SOPRANO under Grant No. 690 724. The computing facilities have been provided by SimTech, funded by the DFG under Germany's Excellence Strategy—EXC 2075 - 390740016.

Author contributions

Basil Jose 0009-0007-7707-8614
Conceptualization (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Software (lead), Visualization (lead), Writing – original draft (lead), Writing – review & editing (equal)

Klaus Peter Geigle 0000-0002-7564-1510
Data curation (equal), Resources (equal), Writing – review & editing (supporting)

Fabian Hampp 0000-0002-9895-4288
Conceptualization (equal), Funding acquisition (lead), Investigation (equal), Project administration (lead), Resources (lead), Supervision (lead), Writing – original draft (supporting), Writing – review & editing (equal)

References

[1] Sun L, Shi J, Wu X, Sun Y and Zeng G 2019 Photon-limited imaging through scattering medium based on deep learning *Opt. Express* **27** 33120–34
[2] Goy A, Arthur K, Li S and Barbastathis G 2018 Low photon count phase retrieval using deep learning *Phys. Rev. Lett.* **121** 243902
[3] Steinberg A M and Driscoll J 2009 Straining and wrinkling processes during turbulence-premixed flame interaction measured using temporally-resolved diagnostics *Combust. Flame* **156** 2285–306
[4] Hampp F and Lindstedt R P 2017 Quantification of combustion regime transitions in premixed turbulent DME flames *Combust. Flame* **182** 248–68

[5] Hampp F, Shariatmadar S and Lindstedt R P 2019 Quantification of low Damköhler number turbulent premixed flames *Proc. Combust. Inst.* **37** 2373–81
[6] Stöhr M, Boxx I and Geigle K P 2023 Flux-based measurement of local scalar source terms in turbulent reacting flows: theory and application to soot formation in an aero-engine model combustor *Combust. Flame* **258** 112456
[7] Hampp F and Lindstedt R P 2017 Strain distribution on material surfaces during combustion regime transitions *Proc. Combust. Inst.* **36** 1911–8
[8] Ranade R and Echekki T 2019 A framework for data-based turbulent combustion closure: a priori validation *Combust. Flame* **206** 490–505
[9] Wang G R and Fiedler H E 2000 On high spatial resolution scalar measurement with LIF part 2: the noise characteristic *Exp. Fluids* **29** 265–74
[10] Brownell C J and Su L K 2008 Planar laser imaging of differential molecular diffusion in gas-phase turbulent jets *Phys. Fluids* **20** 035109
[11] Severin M, Lammel O, Ax H, Lückerath R, Meier W, Aigner M and Heinze J 2018 High momentum jet flames at elevated pressure: Detailed investigation of flame stabilization with simultaneous particle image velocimetry and OH-LIF *J. Eng. Gas Turbines Power* **140** 041508
[12] Böhm B, Heeger C, Boxx I, Meier W and Dreizler A 2009 Time resolved conditional flow field statistics in extinguishing turbulent opposed jet flames using simultaneous high speed PIV/OH PLIF *Proc. Combust. Inst.* **32** 1647–54
[13] Chen H, Abdallah M, Dreizler A, Böhm B and Li T 2024 Simultaneous LII, PAH-LIF, OH-LIF and Mie scattering measurements in solid fuel particle combustion *Proc. Combust. Inst.* **40** 105711
[14] Willert C E and Gharib M 1991 Digital particle image velocimetry *Exp. Fluids* **10** 181–93
[15] Stöhr M, Geigle K P, Hadef R, Boxx I, Carter C D, Grader M and Gerlinger P 2019 Time-resolved study of transient soot formation in an aero-engine model combustor at elevated pressure *Proc. Combust. Inst.* **37** 5421–8
[16] Raffel M, Willert C E, Scarano F, Kähler C J, Wereley S T and Kompenhans J 2018 *Particle Image Velocimetry: A Practical Guide* (Springer)
[17] Geigle K P, Hadef R, Stähr M and Meier W 2017 Flow field characterization of pressurized sooting swirl flames and relation to soot distributions *Proc. Combust. Inst.* **36** 3917–24
[18] Petry N, Schäfer D, Lammel O and Hampp F 2022 Quantification of coflow effects on primary atomization of pressure swirl atomizers *Int. J. Multiphase Flow* **149** 103946
[19] Yin Z, Lammel O, Lückerath R, Ax H and Poulios D 2024 Advancements in wall film thickness measurement under high pressure and high temperature conditions for gas turbine applications *Proc. ASME Turbo Expo* **37** GT2024–123836
[20] Willert C 1996 The fully digital evaluation of photographic PIV recordings *Appl. Sci. Res.* **56** 79–102
[21] Jose B and Hampp F 2024 Machine learning based spray process quantification *Int. J. Multiphase Flow* **172** 104702
[22] Sankaran A, Hain R and Kähler C J 2024 Particle image based simultaneous velocity and particle concentration measurement *Meas. Sci. Technol.* **35** 065206
[23] Picart P, Tankam P and Song Q 2011 Experimental and theoretical investigation of the pixel saturation effect in digital holography *J. Opt. Soc. Am. A* **28** 1262–75
[24] Giepman R H M, Schrijer F F J and van Oudheusden B W 2015 High-resolution PIV measurements of a transitional shock wave–boundary layer interaction *Exp. Fluids* **56** 113
[25] Peterson B, Baum E, Böhm B, Sick V and Dreizler A 2013 High-speed PIV and LIF imaging of temperature stratification in an internal combustion engine *Proc. Combust. Inst.* **34** 3653–60
[26] Zong Y, Brown S W, Johnson B C, Lykke K R and Ohno Y 2006 Simple spectral stray light correction method for array spectroradiometers *Appl. Opt.* **45** 1111–9
[27] Lee J-S, Wee T-L (E) and Brown C M 2014 Calibration of wide-field deconvolution microscopy for quantitative fluorescence imaging *J. Biomol. Tech.* **25** 31–40
[28] Hao Y, Pei H, Lyu Y, Yuan Z, Rizzo J R, Wang Y and Fang Y 2022 Understanding the impact of image quality and distance of objects to object detection performance (arXiv:2209.08237)
[29] Rossi M and Barnkob R 2020 A fast and robust algorithm for general defocusing particle tracking *Meas. Sci. Technol.* **32** 014001
[30] Strässle R M, Faldella F and Doll U 2024 Deep learning-based image segmentation for instantaneous flame front extraction *Exp. Fluids* **65** 94
[31] Touvron H *et al* 2023 LLaMA: Open and efficient foundation language models (arXiv:2302.13971)
[32] Pierson H A and Gashler M S 2017 Deep learning in robotics: a review of recent research *Adv. Robot.* **31** 821–35
[33] Bi K, Xie L, Zhang H, Chen X, Gu X and Tian Q 2023 Accurate medium-range global weather forecasting with 3d neural networks *Nature* **619** 533–8
[34] Huang J, Chai J and Cho S 2020 Deep learning in finance and banking: A literature review and classification *Front. Bus. Res. China* **14** 13
[35] Herrmann B, Baddoo P J, Semaan R, Brunton S L and McKeon B J 2021 Data-driven resolvent analysis *J. Fluid Mech.* **918** A10
[36] Vinuesa R, Brunton S L and McKeon B J 2023 The transformative potential of machine learning for experiments in fluid mechanics *Nat. Rev. Phys.* **5** 536–45
[37] Jose B, Lammel O and Hampp F 2025 ML-based semantic segmentation for quantitative spray atomization description *Int. J. Multiphase Flow* **187** 105179
[38] Hu F, Zhou M, Yan P, Bian K and Dai R 2019 PCANet: a common solution for laser-induced fluorescence spectral classification *IEEE Access* **7** 107129–41
[39] Packard R R S *et al* 2017 Automated segmentation of light-sheet fluorescent imaging to characterize experimental doxorubicin-induced cardiac injury and repair *Sci. Rep.* **7** 8603
[40] Lin T Y *et al* 2014 Microsoft COCO: common objects in context (arXiv:1405.0312)
[41] Li B, Xu Y, Wang Y, Li L and Zhang B 2024 The student-teacher framework guided by self-training and consistency regularization for semi-supervised medical image segmentation *PLoS One* **19** 1–18
[42] Harms T D, Brunton S L and McKeon B J 2024 Estimating dynamic flow features in groups of tracked objects (arXiv:2408.16190)
[43] Kirillov A *et al* 2023 Segment anything (arXiv:2304.02643)
[44] Xue Y, Cheng S, Li Y and Tian L 2019 Reliable deep-learning-based phase imaging with uncertainty quantification *Optica* **6** 618
[45] Liu T, Liu J, Li D and Tan S 2025 Bayesian deep-learning structured illumination microscopy enables reliable super-resolution imaging with uncertainty quantification *Nat. Commun.* **16** 5027
[46] Huang Y *et al* 2024 Segment anything model for medical images? *Med. Image Anal.* **92** 103061

[47] Ma J, He Y, Li F, Han L, You C and Wang B 2024 Segment anything in medical images *Nat. Commun.* **15** 654
[48] Tobin J, Fong R, Ray A, Schneider J, Zaremba W and Abbeel P 2017 Domain randomization for transferring deep neural networks from simulation to the real world (arXiv:1703.06907)
[49] Tremblay J *et al* 2018 Training deep networks with synthetic data: Bridging the reality gap by domain randomization (arXiv:1804.06516)
[50] Toth D, Cimen S, Ceccaldi P, Kurzendorfer T, Rhode K and Mountney P 2019 Training deep networks on domain randomized synthetic x-ray data for cardiac interventions *Proc. 2nd Int. Conf. Med. Imaging Deep Learning* pp 469–82
[51] Tiboni G, Arndt K, Averta G, Kyrki V and Tommasi T 2023 Online vs. offline adaptive domain randomization benchmark (arXiv:2206.14661)
[52] Muratore F, Eilers C, Gienger M and Peters J 2021 Data-efficient domain randomization with bayesian optimization *IEEE Robot. Autom. Lett.* **6** 911–8
[53] Mehta B, Diaz M, Golemo F, Pal C J and Paull L 2019 Active domain randomization (arXiv:1904.04762)
[54] Ruiz N, Schulter S and Chandraker M 2019 Learning to simulate (arXiv:1810.02513)
[55] Litvinov I, Yoon J, Noren C, Stöhr M, Boxx I and Geigle K P 2021 Time-resolved study of mixing and reaction in an aero-engine model combustor at increased pressure *Combust. Flame* **231** 111474
[56] Geigle K P, Köhler M, O'Loughlin W and Meier W 2015 Investigation of soot formation in pressurized swirl flames by laser measurements of temperature, flame structures and soot concentrations *Proc. Combust. Inst.* **35** 3373–80
[57] Grader M, Eberle C and Gerlinger P 2020 Large-eddy simulation and analysis of a sooting lifted turbulent jet flame *Combust. Flame* **215** 458–70
[58] Ho J, Jain A and Abbeel P 2020 Denoising diffusion probabilistic models (arXiv:2006.11239)
[59] Ronneberger O, Fischer P and Brox T 2015 U-Net: convolutional networks for biomedical image segmentation (arXiv:1505.04597)
[60] Wu Y, Kirillov A, Massa F, Lo W Y and Girshick R 2019 Detectron2
[61] He K, Gkioxari G, Dollár P and Girshick R 2020 Mask R-CNN (arXiv:1703.06870)
[62] Lloyd S 1982 Least squares quantization in PCM *IEEE Trans. Inf. Theory* **28** 129–37
[63] Kang Y, Basil J and Hampp F 2025 Pushing fuel flexibility by AM μ-slit injector for high momentum jet stabilised combustion *Proc. 13th Mediterr. Combust. Symp. (MCS13)*