https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

Deep learning-based object detection on LiDAR-derived hillshade images: Insights into grain size distribution and longitudinal sorting of debris flows

Paul E. Schmid[1], Jacob Hirschberg[1,2], Raffaele Spielmann[1,2], and Jordan Aaron[1,2]

[1]Chair of Engineering Geology, Department of Earth and Planetary Sciences, ETH Zürich, Zürich, Switzerland
[2]Swiss Federal Institute for Forest, Snow, and Landscape Research WSL, Birmensdorf, Switzerland

Correspondence: Jacob Hirschberg (jacob.hirschberg@eaps.ethz.ch)

Abstract. Debris flows are hazardous natural phenomena characterized by rapid movements of sediment-water mixtures in steep channels, posing significant risks to life and infrastructure. This study introduces a novel method that leverages hillshade images derived from a high temporal resolution LiDAR scanner and deep learning-based object detection models to analyze debris-flow dynamics. By transforming 3D point clouds into hillshade projections, the method enables efficient detection and tracking of key flow features, including boulders, rolling boulders, surge waves, and woody debris, independent of ambient light conditions. Outputs include object velocities, sizes, and tracks, offering high-resolution insights into debris-flow phenomena such as longitudinal sorting. Six state-of-the-art object detection models were evaluated, with YOLOv11 achieving the best balance of precision, recall, and processing speed. The proposed framework is scalable, significantly reduces processing time compared to manual analysis, and sets the foundation for real-time monitoring and analysis of debris flows across diverse locations and conditions.

1 Introduction

Debris flows are rapid mass movements (>5 m/s) characterized by a mixture of sediment and water, typically occurring in steeply inclined channels (Hungr et al., 2014). Due to their high velocity and surging behavior, they cause thousands of fatalities per year and significantly damage infrastructure, agricultural land, and property (Dowling and Santi, 2014; Prakash et al., 2024; Hilker et al., 2009). The risk posed by debris flows is steadily increasing as development pressures drive construction into areas prone to these hazards (Jakob and Hungr, 2005). Furthermore, in the course of climate change, the occurrence of debris flows is expected to change (Hirschberg et al., 2021b; Kaitna et al., 2023). Due to permafrost degradation and increased rainfall intensities, debris-flow activity in high-alpine areas has increased since the 1950s (Jacquemart et al., 2024). A better understanding of fundamental debris-flow mechanisms is required for future management of these changing boundary conditions. We are currently limited by a lack of in-situ monitoring data that can provide the foundation for further understanding of debris-flow processes. In particular, we lack measurements of the influence of the coupled interaction of individual features (boulders, surge waves, woody debris) and the surrounding liquefied slurry, which are are essential to further our process understanding.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

Theoretical and experimental studies suggest that longitudinal sorting of the constituent materials generates the characteristic steep, unsaturated, bouldery front of debris flows, followed by a liquid slurry (Gray, 2018). As the front progresses more slowly than the liquid portion, large boulders may be deflected or overridden within the flow (Pouliquen et al., 1997; Thornton and Gray, 2008). However, very few in-situ measurements of this process have been made, which limits our ability to understand and predict its important influence on debris-flow motion.

Another important phenomenon associated with debris-flows is the occurrence of surges (Iverson, 1997; Hungr et al., 2014). These surges often show a wave-like appearance and move at a higher velocity than the main flow, which in turn increases impact pressures and enhances their destructive power (Edwards and Gray, 2015). The causes of these surges may include longitudinal sorting of materials (Iverson et al., 2010) or temporary accumulation in gentler channel slopes, followed by sudden release (Kean et al., 2013). They are also frequently explained by mechanisms similar to the formation of water roll waves in open channel flows (Zanuttigh and Lamberti, 2007). As outlined above, the internal processes, the dynamics of individual objects, and the development of surges remain poorly understood, requiring monitoring at higher spatial and temporal resolutions.

Debris flows occur repeatedly within established catchments and known channels (Hungr et al., 2014), which allows for in-situ monitoring of key flow variables. However, the spatial- and temporal resolution of past in-situ measurements has limited our understanding of the interactions between debris-flow features, channel geometry, and flow velocity. Different monitoring systems can capture various aspects of debris-flow dynamics (cf. Hürlimann et al., 2019), but they are often unable to detect a high number of debris-flow features (boulders, surges, etc.) in both day and night conditions. Recently, Aaron et al. (2023) utilized data from multiple high-frequency LiDAR sensors, initially installed in 2019 at the Illgraben Channel (Valais, Switzerland) and supplemented with additional sensors in subsequent years, to estimate debris-flow velocities. However, processing the substantial volume of 3D point cloud data generated during these events (1-2 TB per event) is both time-consuming and labor-intensive. Therefore, an automated method is required to track a greater number of objects more efficiently than is feasible for a human operator.

Convolutional neural networks (CNNs) offer a potential solution for object detection by automating the identification of targets and reducing dependence on manual annotation. While deep learning models like PointNet (Qi et al., 2017) and PIXOR (Yang et al., 2019) have demonstrated effectiveness in processing point cloud data, challenges remain due to the high dimensionality, sparsity, and unstructured nature of such data, which demand significant computational resources and extensive annotation efforts (Fernandes et al., 2021). Aaron et al. (2023) addressed this challenge by taking advantage of the spatial surface structure of debris flows to generate hillshade images from 3D point clouds and applying a particle image velocimetry (PIV) method to the image. Additionally, a LiDAR-camera fusion method was introduced, that leverages optical RGB images from a stationary camera to detect individual features within the debris flow. By combining optical images with point cloud data, valuable 3D information was extracted. Building on this approach, Hirschberg and Aaron (2024) refined the method and evaluated the generalizability of two object detectors. This resulted in high detection accuracies and the development of a multi-domain detector, replacing the need for separate detectors configured to individual locations.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

In this study, we introduce a new method to detect and track debris-flow features based solely on LiDAR data, thereby being independent of ambient light, and enabling the estimation of object velocities, sizes, and trajectories for thousands of debris-flow features. We analyze data captured by multiple high-resolution LiDAR sensors installed at various locations along the channel of a Swiss debris-flow catchment (Illgraben) during a debris-flow event, allowing us to study debris-flow evolution over time and space. To simplify processing, we propose converting point clouds into raster data, followed by hillshade projections. This geospatial representation visualizes relief through shading, enhancing the clarity of 3D visualization. This approach allows us to utilize modern CNN object detection algorithms, originally developed for optical images. Here, we analyze and compare three versions of YOLO (v5, v8, and v11), along with RT-DETR, RetinaNet, and Faster-RCNN for detection and tracking of debris-flow features. This automated approach establishes a foundation for future investigations of debris-flow events across diverse locations and conditions.

2 Study site

2.1 Illgraben catchment

The study site is the Illgraben (46.27° N 7.61° E), a torrent system located in the municipality of Leuk, canton of Valais, Switzerland. The Illgraben is one of the most active debris-flow catchments in the Alps, where hillslope processes deposit sediments in the channel system. These sediments are then mobilized after intense rainfall, developing into debris flows, which can reach the Rhône river (see Fig. 1a) (McArdell and Sartori, 2021). To mitigate damage to the nearby village of Susten, concrete check dams (CDs) have been build along channel (Lichtenhahn, 1971). The Illgraben has been monitored by the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL) since June 2000 (Hürlimann et al., 2003, 2019), with several debris flows recorded annually (Hürlimann et al., 2003; Hirschberg et al., 2021a). Over the years, researchers have deployed a range of sensors to measure diverse debris-flow properties, including front velocity, flow depth, bulk density, normal and shear forces, and erosion depth (e.g., Badoux et al., 2009; Bennett et al., 2013; Berger et al., 2011a, b; Hürlimann et al., 2003; McArdell et al., 2007; McArdell, 2016). Recently, LiDAR scanners were installed 6-10 m above three of the check dams (CDG, CD27, and CD29; see Fig. 1b). The multi-beam LiDAR scanners in place during the event analyzed herein are *OS0* and *OS1* models from *Ouster*, operating at a sampling frequency of 10 Hz. They have a field of view of 33.2 degrees by 360 degrees, made possible by their rotating beam mechanism. While the sensors installed at CDG and CD29 record with a horizontal resolution of 128 channels and 2048 points vertically (a total of 262 144 points), the sensor at CD27 operates with only 64 channels (131 072 points). These sensors record 3D point clouds of moving debris flows, which can then be further analyzed (e.g., Aaron et al., 2023; Spielmann and Aaron, 2024; Åberg et al., 2024).

2.2 5 June 2022 debris flow

In this study, we examine a debris-flow event that occurred on 5 June 2022. The debris-flow front arrived at CDG with a velocity of 5.5 m/s, slowing down to 3.4 m/s at CD27 and to 2.8 m/s at CD29 and the event lasted for around 30 minutes

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

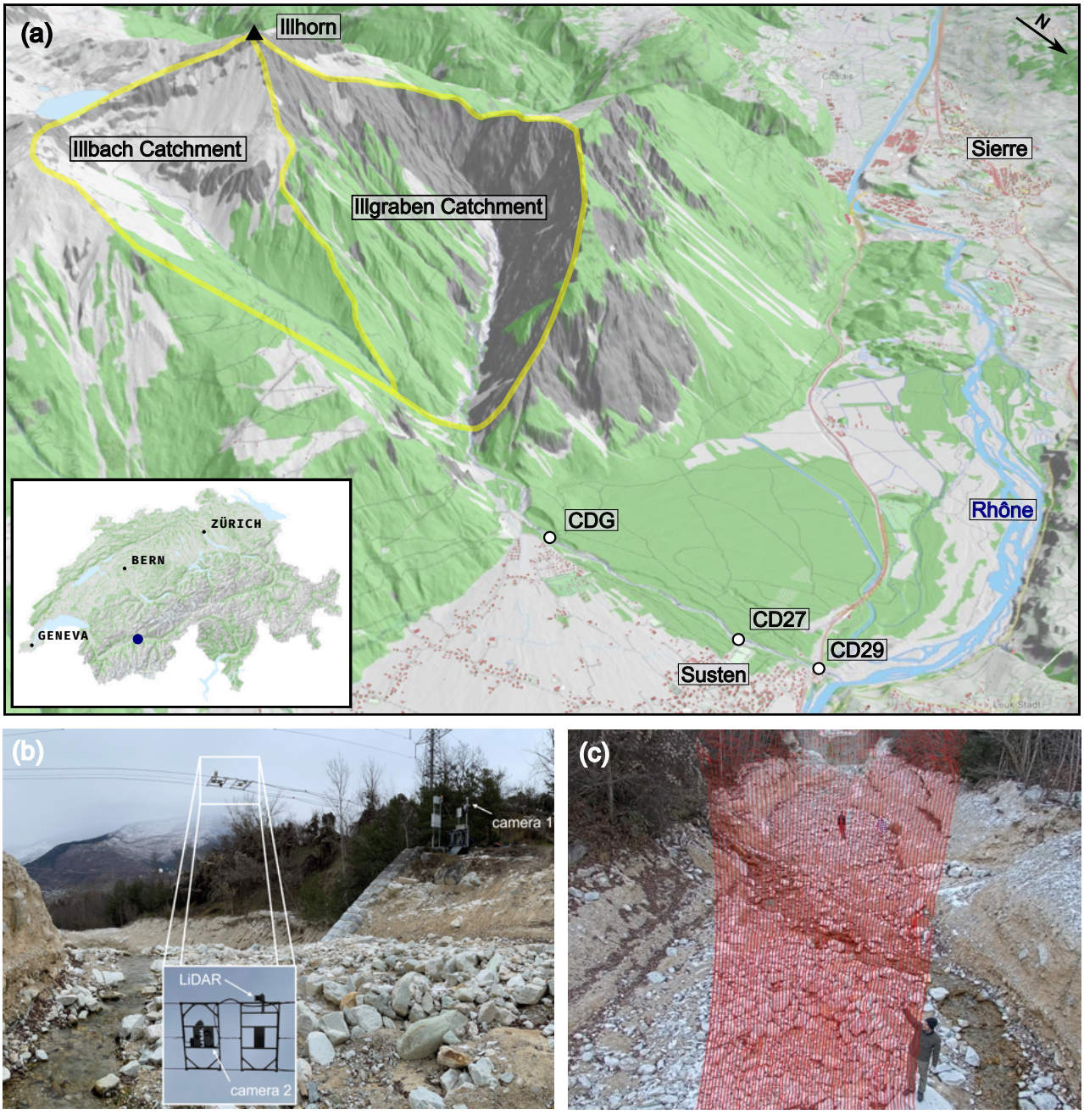

Figure 1. (a) Geographic overview of the Illgraben Channel (46.27° N, 7.61° E) in the Rhone Valley, Switzerland. The catchment boundary is outlined in yellow, and three monitoring sites (CDG, CD27, and CD29) are marked along the Illgraben channel, with inter-site distances of 1 270 m (CDG-CD27) and 470 m (CD27-CD29). 3D base map ©Swisstopo. (b) Overview of the monitoring station (CDG) showing the placement of Camera 1, Camera 2 and a *Ouster* LiDAR Sensor system mounted on a suspended cableway structure (6-10 m above the check dam), facing downstream. (c) Visualization of the LiDAR scan coverage (red points) over the debris-flow channel, facing upstream.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

(Spielmann et al., 2024). However, in this study, we only analyze the first 15 minutes of the event, which includes the majority of visible debris-flow features and flow volume. The event was further characterized by a steep bouldery front (cf. Fig. 3), surge waves and rolling boulders throughout the event. Over 60 surge waves developed in the lower part of the fan such that they were only visible at CD27 and CD29, but not at CDG. Rolling boulders with the size approximately corresponding to the flow depth occurred throughout the event (Hirschberg and Aaron, 2024).

3 Methods

Our study integrates 3D point cloud data from LiDAR sensors and applies CNN models to accurately detect debris-flow features at three monitoring stations along the Illgraben channel, demonstrating the potential of deep learning for hazard monitoring. Figure 2 illustrates a process workflow to detect and track those objects with minimal labeling effort. The results derived from this workflow are compared with velocities obtained through a LiDAR-camera fusion method described by Hirschberg and Aaron (2024), showing that an alternative approach yields comparable results. Additionally, we incorporate velocity data from the same event, calculated using the Particle Image Velocimetry (PIV) method outlined by Aaron et al. (2023) and Spielmann et al. (2024). The PIV method generates dense surface velocity vector fields, which can then be averaged over a specific channel cross-section or rectangular area over time, providing a consistent surface velocity measurement for the selected section (Thielicke and Sonntag, 2021).

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

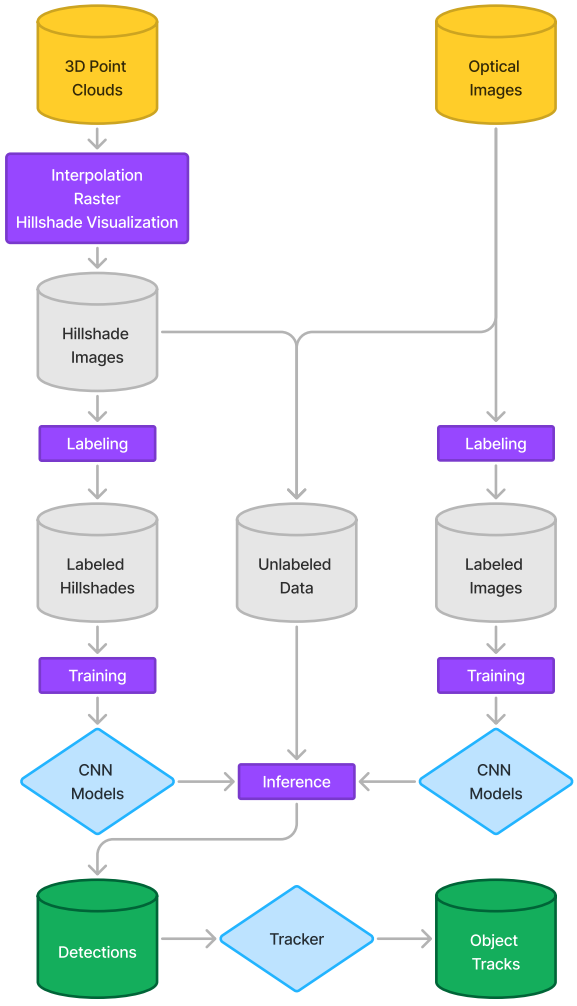

Figure 2. Workflow of the proposed method utilizing hillshade projections (left) and a LiDAR-camera fusion (right) as presented by Hirschberg and Aaron (2024). The process begins with the interpolation, rasterization, and hillshading of 3D point clouds collected from LiDAR sensors. These hillshade images are manually annotated by a human operator. Subsequently, six object detection models are trained using the labeled data. The trained models are applied to unlabeled data during inference, generating object detections and tracks. In the diagram, cylinders represent data inputs, while rectangles denote processing steps.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

3.1 Hillshade image generation

To generate the hillshade images, the raw point clouds were first subjected to pre-processing. The coordinates of the point clouds were transformed into a coordinate system aligned with the channel (x-axis parallel to the channel, y-axis perpendicular, and z-axis vertical). Corrections were applied to account for sensor wobbling caused by wind impacts on the LiDAR installation. Artifacts in the point clouds, such as holes resulting from raindrops on the LiDAR sensor, were removed by selecting a specific area of the point cloud. The point clouds were further cropped to include only the section of the channel upstream of the check dam (Fig. 3b), where the resolution is highest and the results remain unaffected by the presence of the check dam. Additionally, to ensure consistent image dimensions, all point clouds were cropped using the same X and Y limits. For creating the hillshade projections, the point clouds in .PLY format were imported to Python using the Open3D library (Zhou et al., 2018). A mesh was then generated with the aforementioned X and Y limits and a grid spacing of 0.02 m to achieve a consistent 2D hillshade projection. Based on this mesh, the point cloud data were interpolated and subsequently rasterized (Fig. 3c). The resulting raster model was visualized using a hillshading algorithm (Fig. 3d), implemented via the Relief Visualization Toolbox (Kokalj and Somrak, 2019; Kokalj et al., 2022). The azimuth angle was set to 90° and the altitude angle to 45°, as this configuration leads to the highest contrast of the debris-flow topography. The final output consisted of hillshade images with a 2 cm pixel resolution, captured at intervals of 0.1 s.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

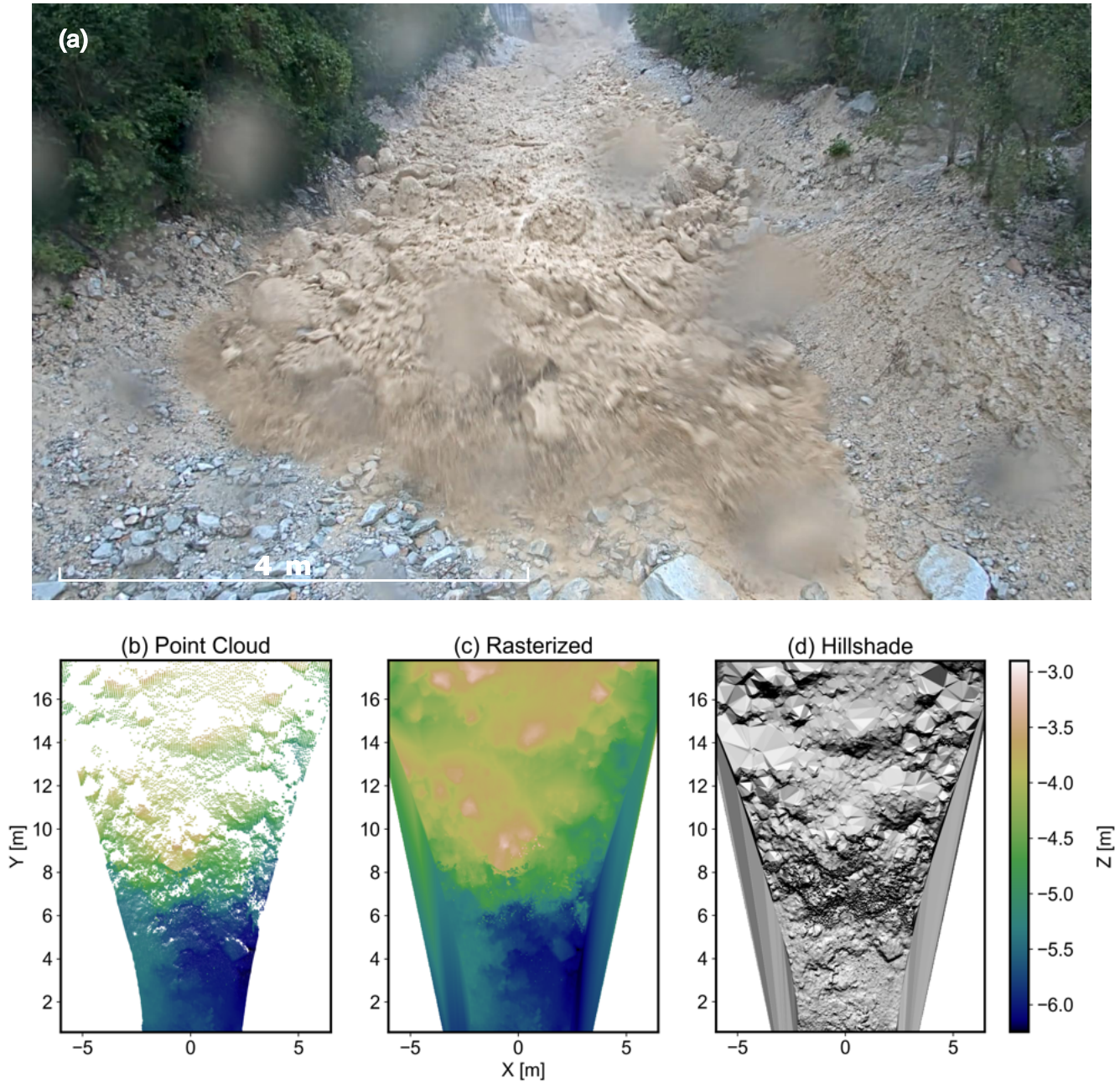

Figure 3. Multi-representation of debris-flow data. (a) Image capture (Camera 2) of an active debris-flow event (5 June 2022) in the Illgraben Channel, showcasing the flow front. (b) The corresponding 3D point cloud representation generated by the *Ouster OS1* LiDAR sensor illustrates the spatial distribution of points. (c) Rasterized and interpolated elevation model derived from the point cloud, providing a gridded representation of the debris-flow surface. (d) Hillshade visualization, emphasizing terrain features through simulated shading for enhanced interpretation of morphological characteristics.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

3.2 Labeled dataset of hillshade projections

To ensure a consistently labeled dataset, all annotations were created by a single individual under the supervision of experts with substantial knowledge of the domain and study site. The open-source data labeling platform, Label Studio (HumanSignal, 2024), was used for this task. Additional video and camera material were used as a reference during the labeling process to accurately identify debris-flow objects. When objects were considered significant enough to be classified into one of the four categories, a rectangular box was drawn around the object of interest, referred to as a label in subsequent discussions. Significance was defined as the ability to distinguish the object from the background noise in the hillshade image. The objects of interest were defined and tested in previous studies by Aaron et al. (2023); Spielmann and Aaron (2024); Hirschberg and Aaron (2024), and these definitions were adopted in this work. The four feature classes include Boulders, Surge Waves, Rolling Boulders, and Woody Debris (Fig. 4). These objects are hypothesized to represent different vertical velocity profiles at varying depths within the debris flow. For example, woody debris is thought to travel at the surface velocity, while rolling boulders, maintaining a consistent size throughout the flow depth, may reflect a uniform velocity profile across the entire depth (Aaron et al., 2023). To further investigate these properties, we retained the same feature classes. Hillshade images for model training were randomly selected from all three monitoring stations and two events (5 June 2022, 19 September 2021), with additional images included for more complex scenarios, such as the debris-flow front. A total of 1 092 images were selected, with 21 698 instances annotated (Tab. 1). The entire dataset was then divided into training, validation, and testing sets with a ratio of 70:15:15.

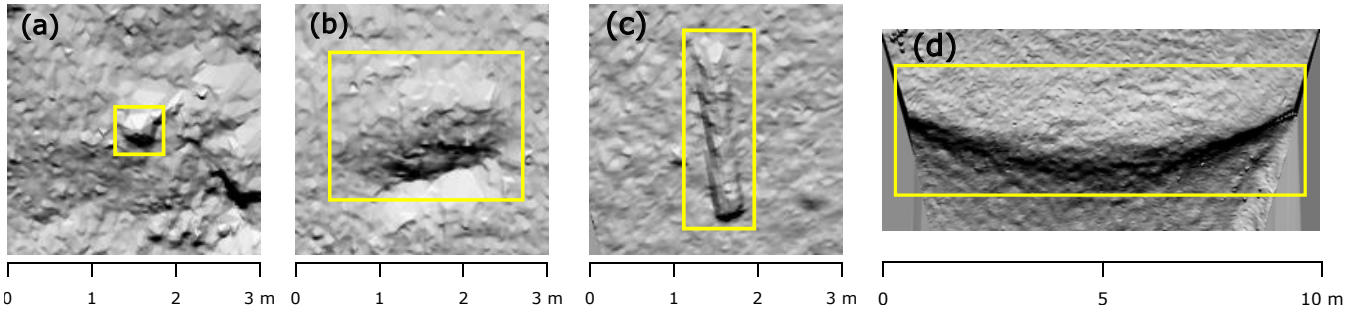

Figure 4. Example annotations of the four feature classes used in the object detection training set. (a) Boulder: small to medium-sized sliding rocks. (b) Rolling Boulder: a displaced rock structure indicating rolling movement, with characteristic turbulence around the object. (c) Wood: an elongated object resembling a log, occasionally including branches or roots. (d) Surge Wave: a surge traveling through the channel over the entire channel width.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

Table 1. Summary of our dataset, detailing the number of images and annotated instances across the training, validation, and test sets for all classes combined (Boulder, Surge Wave, Rolling Boulder and Wood). The dataset was split with a ratio of 70:15:15.

Set	Class	# Images	# Instances
Training	all	764	14 914
Validation	all	164	3 308
Test	all	164	3 476

3.3 Object detection and tracking

We tested and compared four state-of-the-art object detection models, including three versions of YOLO (v5, v8, and v11), as well as RT-DETR, RetinaNet, and Faster R-CNN, using our custom hillshade dataset. The models were selected based on their popularity and performance, all ranking among the top performers in the COCO benchmark (cf. Zou et al., 2023). The largest pre-trained versions of each YOLO model (YOLOv5-X, YOLOv8-X, YOLOv11-X) and RT-DETR-X were deployed, along with a pre-trained ResNet-50 backbone for RetinaNet50 and a ResNeXt-101 backbone for Faster-RCNNX101. The evaluation was conducted using the same dataset, with identical training, validation, and test splits across all models. Training and detection processes were conducted on NVIDIA GeForce RTX 4090 GPUs available through the Euler IX cluster at ETH Zürich.

3.3.1 Object detection models

Detectron2 is a highly modular framework developed by Meta AI Research for state-of-the-art object detection and segmentation tasks. Built on PyTorch, it provides powerful tools for a wide range of applications and features an extensive model zoo with pre-trained weights. Using this framework, we evaluated variations of the RetinaNet (ResNet-50 and ResNet-101) and Faster R-CNN models (ResNet-50, ResNet-101 and ResNeXt-101), where 50 and 101 denote the depth of the backbone network (Wu et al., 2019).

Faster R-CNN is a widely adopted two-stage object detection model that builds upon the success of its predecessors, R-CNN and Fast R-CNN. The key innovation of Faster R-CNN is the introduction of the Region Proposal Network (RPN), which efficiently generates region proposals as part of the model, significantly speeding up the detection process. In the first stage, the RPN proposes candidate object regions, and in the second stage, these proposals are refined and classified into object categories. The model's ResNet backbone is responsible for feature extraction from the input image, while the RPN and the detection head handle proposal generation and final classification, respectively (Ren et al., 2016).

RetinaNet is a widely used one-stage object detector, developed when two-stage detectors within the region-based CNN framework were dominant. The key difference between one-stage and two-stage detectors is that one-stage detectors predict objects and their bounding boxes directly from the image in a single pass, whereas two-stage detectors first generate region-based proposals and then refine them in a second stage. RetinaNet addresses the problem of class imbalance by introducing Focal Loss, a function that reduces the weight of well-classified examples, allowing the model to focus on harder-to-classify,

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

minority class examples without sacrificing performance on the majority class. The model also consists of a ResNet-based backbone (FPN) and two task-specific subnetworks: one for object classification and the other for bounding box regression (Lin et al., 2018).

Building on the advancements in one-stage detectors like RetinaNet, YOLO (You Only Look Once), originally introduced by Redmon et al. (2016), optimized detection speed and accuracy through its unified detection architecture. YOLOv5 (Jocher, 2020) further improved upon this by introducing multi-scale detection, data augmentation, and AutoAnchor algorithm, which search for ill-fitted anchor boxes. It features the CSPDarknet53 backbone and FPN for feature extraction, balancing accuracy and speed effectively (Terven et al., 2023). YOLOv8 (Jocher et al., 2023) builds on YOLOv5 by introducing the C2f module for improved detection accuracy through better feature-context integration. It adopts an anchor-free design with a decoupled head for objectness, classification, and regression tasks, enhancing precision (Terven et al., 2023). The latest YOLO version, YOLOv11, incorporates new architectural elements like the C3k2 and partial Pyramid Pooling - Fast (SPPF) block. The C3k2 module improves feature representation by deepening the network's capacity to capture intricate patterns. Meanwhile, the SPPF block accelerates the model by pooling features at multiple scales, enabling better handling of objects of varying sizes with minimal computational overhead Jocher and Qiu (2024).

Most recent object detectors, such as the YOLO series, use a Non-Maximum Suppression (NMS) algorithm to filter overlapping predictions, such as bounding boxes in object detection. NMS removes predictions below a probability threshold and selects the highest-probability entity, discarding others with an Intersection over Union (IoU) above a specific threshold (typically 0.5). However, NMS slows down inference and introduces hyperparameters that destabilize both speed and accuracy (Zhao et al., 2024). Recently, a new object detection model was introduced by Zhao et al. (2024), presenting a Transformer-based detector. The authors suggest that their Real-Time Detection Transformer (RT-DETR) can address these issues. Specifically, by separating intra-scale interactions (features within the same scale) and cross-scale fusion (features from multiple scales) to enhance speed, an effective hybrid encoder is developed that efficiently processes multi-scale data. The uncertainty-minimal query selection, which minimizes the discrepancy between classification and localization confidence, is proposed to provide the decoder with high-quality initial queries, thereby improving accuracy (Zhao et al., 2024).

For the evaluation of the trained models, we employed the widely used object detection metrics: Precision (P), Recall (R), F1-Score, and mean Average Precision (mAP). P represents the accuracy of positive predictions, while R reflects the model's sensitivity. The F1-Score is the harmonic mean of P and R. The mAP is calculated by integrating the P-R curve across various Intersection over Union (IoU) thresholds (commonly 0.5 and 0.5 to 0.95). A more detailed definition of these metrics can be found in the Appendix A1.

3.3.2 Data augmentation

The choice of data augmentation techniques can significantly impact model performance and robustness (Shorten and Khoshgoftaar, 2019). To address class imbalance, we implemented a basic data augmentation strategy by horizontally flipping images of underrepresented classes. The feature class boulder accounts for the majority of instances in the training dataset, with 13 882 occurrences, while surge wave, rolling boulder, and wood are represented by only 73, 459, and 500 instances, respectively. To

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

mitigate this imbalance, we selected images containing one or more of these underrepresented classes and reintegrated them into the training dataset. This augmentation approach was consistently applied across all tested models. Certain models, particularly those optimized for fine-tuning on small amounts of task-specific data, rely on extensive data augmentation to enhance their performance (e.g., YOLOv5). In contrast, most popular implementations of Transformer-based backbones, such as those in Detectron2, provide a default set of augmentations designed for large-scale pretraining, typically limited to random flips, resizing, and cropping (Ruis et al., 2024). For this study, we applied augmentations which were meaningful to our application, including translation, mosaic, and resizing, while omitting augmentations like RGB saturation due to the use of grayscale images. These choices align with the default augmentations recommended by the respective model developers. Specifically, we employed techniques such as Mosaic and Mixup for YOLO models, and image resizing for Detectron2-based models (Tab. 2).

Table 2. Summary of image augmentation techniques and their corresponding parameters applied across all models, including RT-DETR, YOLO versions, and the Detectron2 models Faster-RCNN and RetinaNet. Horizontal flipping was applied exclusively to images containing underrepresented feature classes (Surge Wave, Rolling Boulder, and Wood).

Model	Augmentation	Value/Probability
All models:	Horizontal Flipping	0.5
YOLOv5, YOLOv8,	Translate	0.1
YOLOv11, RT-DETR:	Scale	0.5
	Mosaic	1
	Mixup	0.5
Faster-RCNN, RetinaNet:	Resize Shortest Edge	(640, 672, 704, 736, 768, 800)
	Scale Jittering	0.5

3.3.3 Multi-object tracking

After detecting debris-flow features on individual hillshade images, we employed two Multi-Object-Tracking (MOT) algorithm to unify the detections across subsequent frames. The Simple Online and Realtime Tracking (SORT) algorithm (Bewley et al., 2016; Bewley, 2024) predicts the movement of detected objects between frames using a linear constant velocity model and refines these predictions with a Kalman Filter. To associate predicted positions with actual detections, SORT utilizes the Hungarian algorithm, optimizing the assignment based on an IoU metric. Each object is assigned a unique identifier, which is maintained across frames if the association is successfully established. However, preliminary testing showed that the SORT algorithm performed poorly for objects with velocities greater than 5 m/s. To overcome this limitation, we applied the BoT-SORT algorithm (Aharon et al., 2022). BoT-SORT builds on the tracking-by-detection approach used in SORT but incorporates an improved Kalman filter state vector for more accurate box localization and a camera motion compensation feature.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

3.4 Object velocity

The resulting tracks from the MOT algorithms were used to calculate object velocities. First, we converted the bounding box coordinates from pixel units to meters using the hillshade image resolution of 0.02 m/px. The velocity was then calculated based on the temporal displacement between consecutive frames (Eq. 1). For object tracks that experienced a class change, we assumed that the most frequently occurring class corresponded to the ground truth.

$$v(t_i) = \frac{\sqrt{(x(t_i) - x(t_{i-1}))^2 + (y(t_i) - y(t_{i-1}))^2}}{t_i - t_{i-1}} \quad (1)$$

where v_{t_i} is the velocity of a tracked debris-flow feature and x and y are the object coordinates at time t . The time step between two frames is 0.1 s, corresponding to the LiDAR recording frequency of 10 Hz.

4 Results

4.1 Metrics and model assessment

Table 3 presents a comparison of P, R, F1-score, and mAP at IoU thresholds 0.5 (mAP50) and 0.5-0.95 (mAP50-95) across the four feature classes (cf. Fig. 4) on the test dataset. Overall, YOLOv8-X stands out in precision, while YOLOv11-X tops in mAP50-95, indicating best performance across varying IoU thresholds. Notably, differences in metrics are minor, with variations of $<$0.067 in mAP and $<$0.054 in F1-score across all models. However, the class-averaged metrics are disproportionately influenced by the high precision values of the surge wave class (e.g., P=1 and R=0.988 with the YOLOv5-X model). These trends are also evident when analyzing the metrics by class (see Fig. A1, Tab. A1). The model comparison produced results consistent with the previous observations, with RetinaNet50 showing the lowest performance (particularly for the Boulder class, with mAP50-95=0.382) and YOLOv11-X demonstrating the best performance, showing strong mAP50-95 values across all classes. When analyzing the performance metrics across varying confidence thresholds (CT), it is observed that YOLOv5-X, YOLOv8-X, and YOLOv11-X achieve high F1 scores ($>$0.8) at a low CT value of 0.1 but exhibit a sharp decline when the CT exceeds 0.7 (Fig. 5). In contrast, RT-DETR-X, Faster-RCNN, and RetinaNet reach their peak performance at higher CT values (0.6-0.7) but demonstrate poor precision and recall at lower CT thresholds.

Inference was tested on the hillshade projections for the station CDG using a confidence threshold of 0.1 and an IoU of 0.5 (Tab. 3). In terms of processing speed, the models show considerable differences, ranging from 16.1 frames per second (FPS) (RetinaNet50) to 38.2 FPS (YOLOv11-X). Notably, YOLOv11-X stands out with 38.2 FPS, nearly twice as fast as the next fastest model, YOLOv8-X, with 21.7 FPS. All other models (YOLOv5-X, YOLOv8-X, RT-DETR-X, Faster-RCNNX101, and RetinaNet50) fall within a similar range of 16.1 to 21.7 FPS. Based on its precision and processing speed, the YOLOv11-X model was selected to generate the results for the following analyses.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

Table 3. Performance metrics for object detection of debris-flow features (Boulder, Surge Wave, Rolling Boulder, and Wood) on the test dataset. Metrics include Precision (P), Recall (R), F1 score (F1), mean average precision at an intersection over union threshold of 0.5 (mAP50), and across thresholds from 0.5 to 0.95 (mAP50-90). A confidence threshold of 0.1 was applied. Class-specific metrics are listed in the Appendix Tab. A1. The inference time represents the total duration required to detect debris-flow objects across 17 999 hillshade projections (corresponding to one LiDAR data package and 30 minutes recording material) and to and save the prediction results to a file for each projection. However, for this specific event, analyzing only the first 15 minutes is sufficient because it encompasses the majority of visible debris-flow features and flow volume. FPS (Frames Per Second) describes the number of frames a model can process in one second, indicating its inference speed and real-time capability.

Model	P	R	F1	mAP50	mAP50-95	Inf. Time [s]	FPS
YOLOv5-X	0.866	0.792	0.827	0.859	0.579	1 053	17.1
YOLOv8-X	0.890	0.769	0.825	0.859	0.576	829	21.7
YOLOv11-X	0.817	0.823	0.820	0.862	0.583	471	38.2
RT-DETR-X	0.833	0.799	0.816	0.851	0.545	1 065	16.9
Faster-RCNNX101	0.848	0.796	0.821	0.828	0.519	1 017	17.7
RetinaNet50	0.794	0.753	0.773	0.795	0.506	1 117	16.1

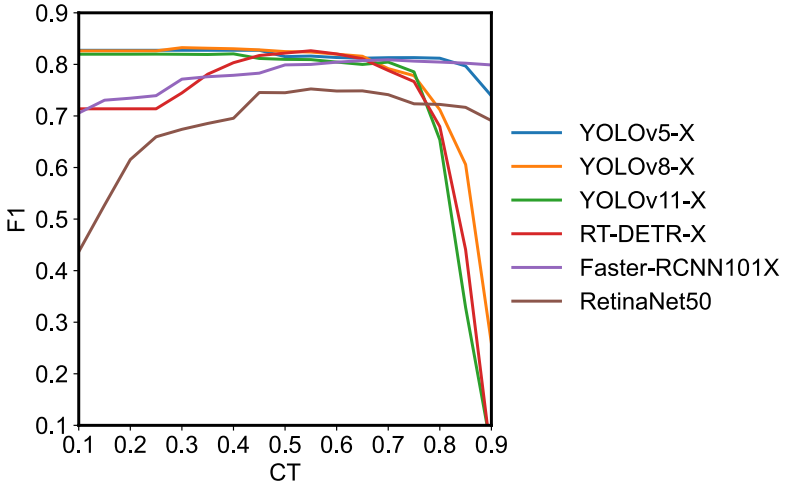

Figure 5. Evaluation of F1 Score sensitivity to Confidence Threshold (CT) for YOLOv5-X, YOLOv8-X, YOLOv11-X, RT-DETR-X, Faster-RCNN101X and RetinaNet50

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

4.2 Object analysis

4.2.1 Tracks and velocities

The computation of object velocity using frame-by-frame tracking with BoT-SORT, provides a detailed dataset of object velocities for each timestep and detected object. The velocities were analyzed either as instantaneous or track-average velocities. Object tracks with an average velocity below 1 m/s were omitted from the final dataset, as their low velocity indicates (temporary) deposition. A total of 23 120 object tracks were recorded and reduced to 16 452 after applying the described filtering (Tab. 4). The number of detected objects (after filtering) varied among the three stations. At CDG, 8 191 objects were detected, with their velocities calculated, while only 5 882 objects were analyzed at CD27 and 2 379 at CD29 for the hillshade method (Tab. 4). As a comparison, in the same table, we also show the number of detected objects on camera images (LiDAR-camera fusion). Interestingly, the number of detections shows a different trend among the stations. Reasons for these observations will be provided in the Discussion.

Table 4. Summary of detections and object counts for all stations (CDG, CD27, and CD29). It includes the total number of detections, the number of unique objects identified on hillshade projections, the number of filtered hillshade objects (after post-processing), and the number of objects identified from the LiDAR-camera fusion method.

Station	Class	# Detections	# Objects LiDAR	# Objects LiDAR (filtered)	# Objects Camera
CDG	all	262 405	12 037	8 191	523
CD27	all	142 298	7 959	5 882	1 005
CD29	all	98 713	3 124	2 379	990
Total	all	503 416	23 120	16 452	2 518

In Figure 6, these object (track-average) velocities are compared to the results from the PIV hillshade surface velocity measurement method. At CDG, the velocities of objects detected by LiDAR were consistently slightly higher compared to the PIV velocities, whereas at CD27 and CD29, both velocity measurements showed good agreement. Additionally, at CD27 and CD29, the LiDAR velocities exhibited little variation and closely followed the PIV velocities. In contrast, at CDG, the LiDAR velocities displayed significant variability, with differences of up to 2 m/s within the same timestep. Overall, the velocities showed a gradual downstream decrease from CDG to CD29. As shown in Figure 6, boulders were predominantly detected at the beginning of the event, while rolling boulders peaked later and persisted towards the end, especially for CDG and CD27. Woody debris, although less frequently detected, appeared consistently throughout the event at all stations. The agreement between object velocities obtained using the LiDAR method and the LiDAR-camera fusion is illustrated in Fig. A2.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

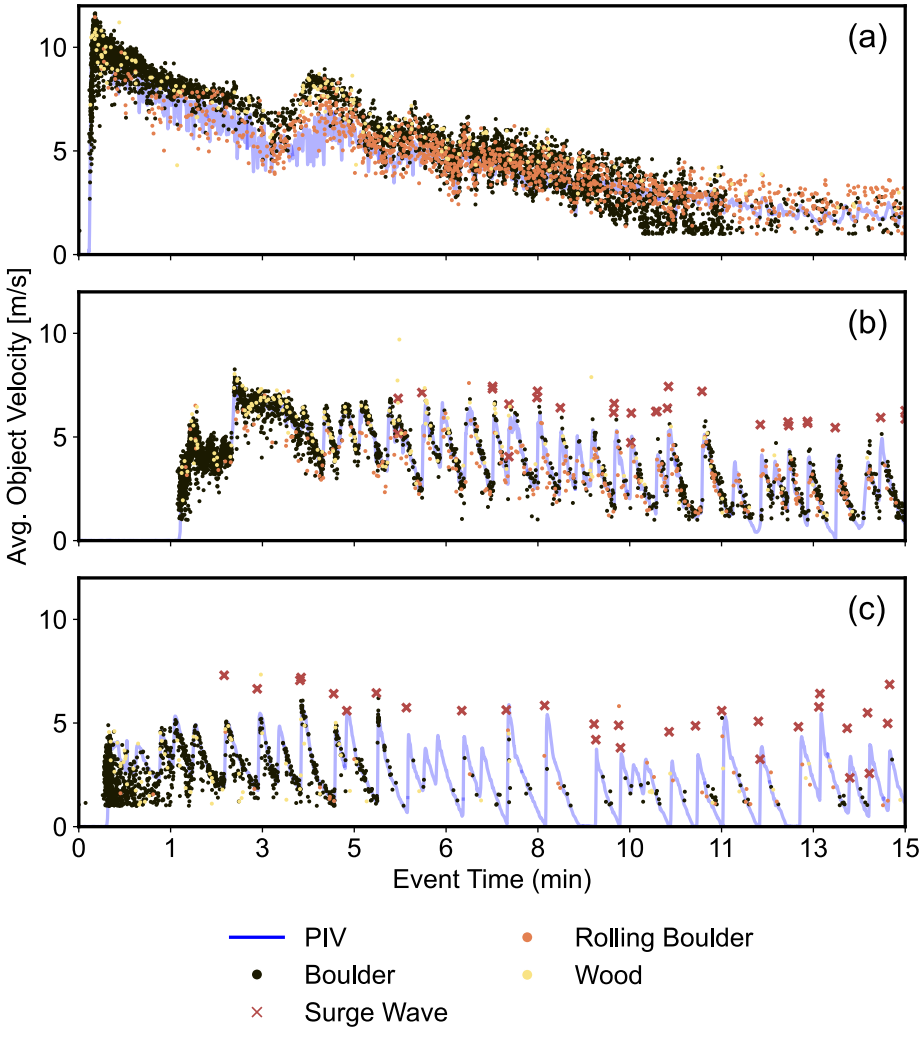

Figure 6. Mean object velocities calculated for the 5 June 2022 event, using the LiDAR hillshade method at the three monitoring stations: (a) CDG, (b) CD27, and (c) CD29. Colors represent the different object classes. Surface velocities derived from particle image velocimetry (PIV) by Spielmann et al. (2024) are included as a reference.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

4.2.2 Sizes and flow depth

Object size comparisons relative to flow depth at each station highlighted notable differences, as indicated in Fig. 7. At CDG, a substantial proportion of tracked objects was larger than the flow depth, while at CD29, nearly all tracked objects were smaller than the flow depth (Fig. 7). Furthermore, the absolute size of the objects shows a decreasing trend from CDG to CD29, with many objects exceeding 2 m at CDG and almost all objects being smaller than 2 m at CD29. Additionally, the ground truth object sizes showed strong correlations with the detected object sizes (SI Fig. S2). The number of tracked objects per station also decreased downstream from CDG to CD29, as illustrated in Fig. 6, 7, and Tab. 4.

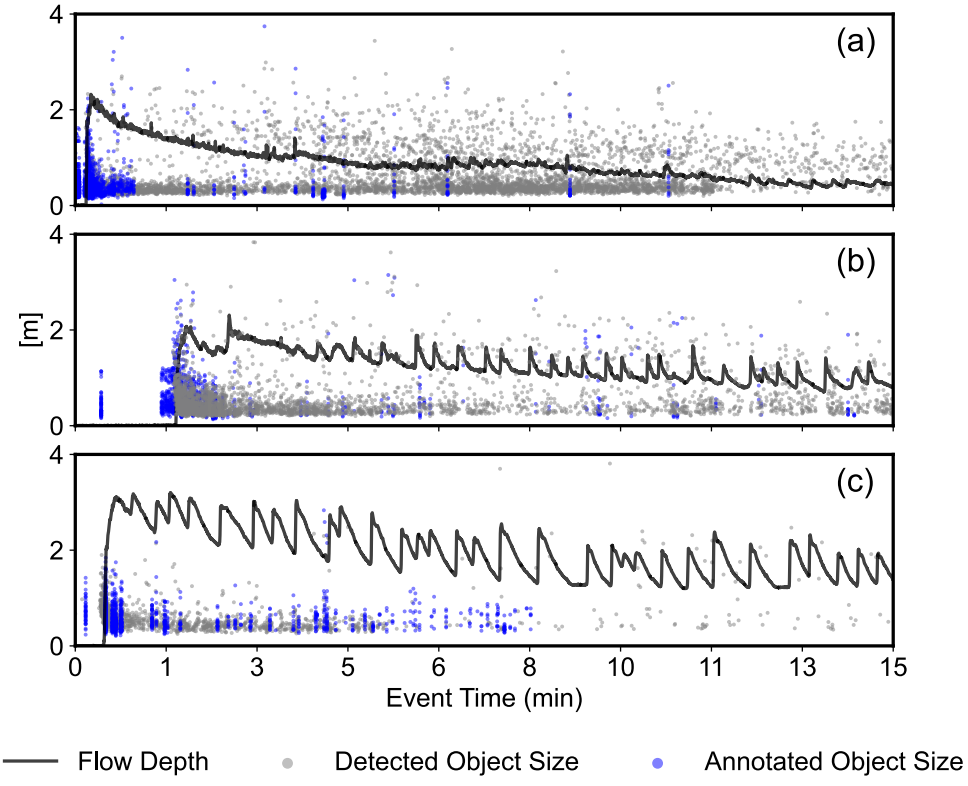

Figure 7. Comparison of flow depth and object size over time at the three monitoring stations (CDG, CD27, and CD29) during the 5 June 2022 event. The black line represents the flow depth (in meters) as a function of event time (in minutes), while the gray markers show LiDAR-detected object sizes. Annotated object sizes (= ground truth bounding box size) are represented by blue markers. The data highlight the temporal evolution of flow dynamics, with larger objects typically observed at CDG and a reduction in object size relative to flow depth downstream at CD29.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

4.3 Grain size distribution and longitudinal sorting

As indicated in Figure 7, the observed boulder and rolling boulder sizes vary over time and location. This allows for studying dynamic grain size distributions (GSDs) along the flow path, as summarized in Fig. 8, where LiDAR detections (Fig. 8a-c) represent all target sizes in a given frame (prior to object tracking), while LiDAR tracks (Fig. 8d-f) indicate the size of individual objects across multiple frames (after unification by the tracking algorithm). As described in the methods, this serves as a filter, reducing the number of observations while enabling the analysis of distinct objects. Additionally, Fig. A4 shows a correlation between the detected object sizes and their ground truth sizes.

At station CDG, the GSD becomes coarser over time (Fig. 8a/d), whereas at station CD29, the trend is reversed, with finer grain sizes observed as time progresses (Fig. 8c/f). Station CD27, located between CDG and CD29, exhibits a less pronounced trend of coarsening over time (Fig. 8b/e). The trends of coarsening at CDG, fining at CD29, and a combination of fining and coarsening at CD27 are evident in both the detection and tracking data. The particle sizes at stations CDG and CD27 (see Tab. A2) are comparable across both the detection and object track datasets, with D50 values ranging from 0.36 to 0.40 m at CDG and 0.39 m at CD27. In contrast, particle sizes at station CD29 are generally larger than those at CDG and CD27, with D50 values ranging from 0.46 to 0.50 m. At stations CD27 and CD29, the median particle size (D50), where 50% of the particles are finer and 50% are coarser, becomes progressively finer over time. In contrast, D50 at station CDG shows minimal change. Notably, at CDG, the D90 (the size where 90% of particles are finer) becomes coarser over time, likely due to the increased presence of rolling boulders after event minute 5 (cf. Fig. 6), a process less prominent at CD27 and CD29. The observed GSDs also allow for the computation of a Uniformity Coefficient (Cu), commonly used in soil characterization. At stations CDG, CD27, and CD29, Cu values range between 1.7 and 1.9. With values below the threshold of 4, the sediment can be classified as poorly graded (BSI, 2020), and there is no considerable difference between the stations.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

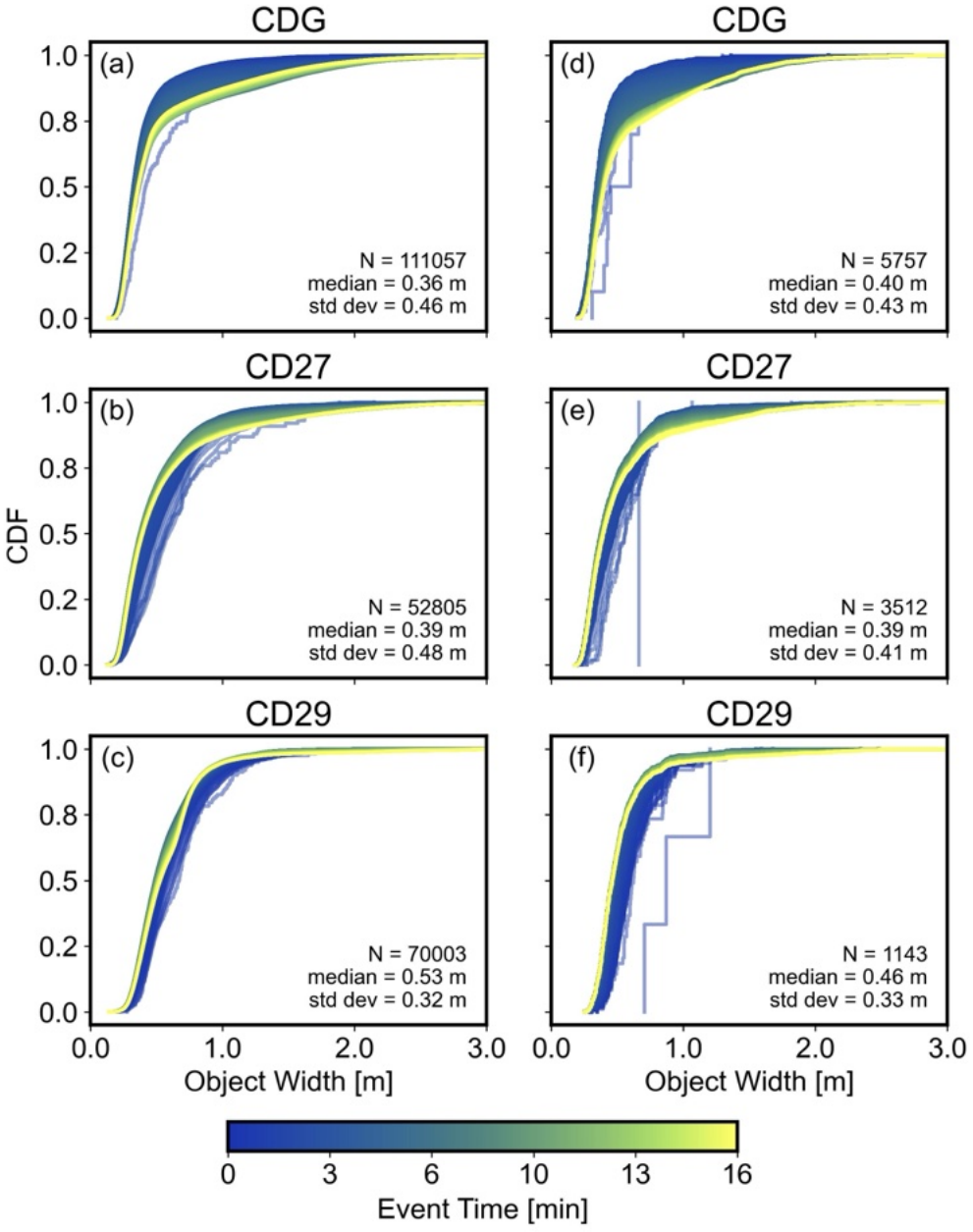

Figure 8. Grain size distribution (GSD) across stations CDG, CD27, and CD29. Cumulative distribution functions (CDFs) of object widths derived from LiDAR detections (a-c) and LiDAR tracks (d-f) are shown for CDG, CD27, and CD29. CDFs are colored by event time to show temporal variations, and each line represents the CDF of all detected and tracked objects until the current event time, respectively. Event start time is defined by the arrival of the debris-flow front at the check dam.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

5 Discussion

5.1 Model performances

We present a novel automated method to extract relevant information on debris-flow features from high-resolution 3D LiDAR measurements at three monitoring stations. By converting raw LiDAR point clouds into hillshade projections and applying object detection algorithms to these 2D images, our approach enables the accurate identification of debris-flow objects. For the debris flow analyzed in this study, about 10 000 objects were automatically identified, which would take a human operator 22 weeks to extract (assuming 10 000 objects, 50 frames per object, 20 s annotation time). The detected objects can then be linked to reconstruct tracks of boulders, woody debris, and surge waves, allowing for the calculation of their velocities. This framework offers a foundation for efficient data processing at other locations and facilitates the analysis of various debris-flow events across diverse settings. This framework is not only transferable but also independent of ambient light conditions, allowing it to be used both day and night. Additionally, it is significantly more efficient than working directly with 3D data.

We tested six different object detectors on their ability to detect debris-flow features. The performance metrics of the different models showed very similar results (Tab. 3) and points to that all of the tested model architectures are equally suited for the debris-flow object detection. The small variations in model accuracy suggest that the limiting factor is not the detector architecture but rather the quality and size of the dataset. The inferences speed, however, reveals major differences between models up a factor of ~ 2.5. Important considerations when selecting a model are therefore the computational and speed requirements, especially for real-time applications. Single-shot detectors, especially YOLOv11, confirmed to be desirable in this regard (Redmon et al., 2016).

The resulting precision and recall metrics, both approximately 0.8, are generally consistent with results reported in other applications, where values between 0.7 and 0.9 are considered good (Diwan et al., 2023). The slightly lower performance may be attributed to the large number of small objects detected, which are inherently more challenging to identify. This difficulty arises because CNN architectures downsample feature maps across several levels, resulting in a loss of spatial information at smaller scales (Liu et al., 2021; Li et al., 2020). Consequently, the precision values for the boulder class, the smallest among the four classes, were notably lower (SI Fig. S3, SI Tab. S1). Nonetheless, the overall metrics were strong, even for smaller classes. However, these results might be slightly influenced by correlations between the datasets, as the images (or frames) were randomly selected. This randomness could result in similar images appearing in different splits (e.g., Frame 1110 in the training set and Frame 1120 in the test set, captured one second apart). Recall values (0.75-0.82) show that many features are missed in the automatic detection, further supporting this. However, precision values are generally high, meaning false classification is low. Some false detections in the background were observed, but these were effectively filtered out by the tracking algorithm. As a result, these errors did not propagate into the calculations of velocities and sizes, nor did they affect their interpretation.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

5.2 Methodological strengths and limitations

The object velocities derived from feature tracking were compared with velocities computed for the same event using PIV by Aaron et al. (2023), showing strong overall agreement across all three monitoring stations (cf. Fig. 6). However, at station CDG, during event minutes 1 to 5, the estimated object velocities for the boulder class are approximately 1-2 m/s higher than the PIV velocities, whereas the rolling boulder class velocities closely match the PIV values. Higher peak velocities were also observed at CD27 and CD29 (Fig. 6b-c) compared to the PIV velocities, particularly for the surge wave class. This discrepancy is expected, as PIV calculates the mean velocity within an 2×2 m box of the velocity field, while object tracking directly follows the actual movement of individual features. The comparison of PIV velocity validates the computation of object velocities based on the center points of detection bounding boxes. Furthermore, the computed velocities align closely with surface velocities (cf. Fig. A2) obtained through the LiDAR-camera fusion method by Hirschberg and Aaron (2024).

Stationary objects in the channel prior to the debris-flow front and objects temporarily deposited on the channel banks were filtered, with re-entrainment by subsequent surge waves accounted for in the analysis. It is important to note that when interpreting the velocity results, the limited field of view of the LiDAR sensors (cf. Figure 1c) restricts velocity measurements to the area covered by the sensors, rather than the entire channel width. This will be addressed by installing additional sensors. Another challenge arises from the conversion of 3D LiDAR point clouds into 2D hillshade images, which leads to the loss of spatial information along the vertical (Z) axis. As a result, the computed object velocities are constrained to a 2D plane, despite the fact that debris objects may submerge and reemerge, indicating vertical velocity components that remain uncaptured.

Despite its effectiveness, this object detection approach has some limitations. Detection errors primarily result from background noise, likely caused by variations in triangulation across sequential frames when using the hillshading algorithm, as well as misclassifications between feature classes. Additionally, the generalizability of the models requires further evaluation, necessitating the expansion of the dataset to include more events and locations. Furthermore, object collisions may lead to tracking errors, where the tracking algorithm incorrectly associates a new object post-collision, introducing further uncertainties in the velocity analysis. By addressing these limitations through expanded datasets, additional LiDAR sensors, and improved detections models, the methodology has the potential to deliver more robust insights into debris-flow dynamics across diverse conditions.

5.3 Implications for debris-flow mechanics

Our proposed method demonstrates the potential to observe GSDs and longitudinal sorting in debris flows to some extent. The bounding boxes of detected objects show a strong correlation with their actual sizes (Fig. A4), supporting the reliability of GSD analysis. A slight offset is observed, likely due to the labeling process, where ground truth annotations were intentionally drawn slightly larger than the actual objects to account for surrounding water turbulence (cf. Figure 4b).

Preliminary analyses of longitudinal sediment sorting indicate temporal variations at each monitoring station, as reflected in the GSDs. These distributions also vary spatially between stations, suggesting changes in flow evolution along the channel. At CDG, minimal variation in D50 (median grain size) indicates that the overall GSD remains relatively stable, with coarsening

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

occurring primarily in the larger particle fraction. In contrast, at CD29, the GSD becomes progressively finer over time, as evidenced by decreasing D50 values. This trend may result from sediment sorting processes, wherein coarser particles are preferentially transported to the flow front, leading to an extended fine-grained trailing slurry. At CD27, located between CDG and CD29, both coarsening and fining occur over time. This mixed trend may reflect its transitional position, where sediment sorting processes are less pronounced compared to the upstream (CDG) and downstream (CD29) stations.

It is important to note that these findings are based on the analysis of hillshade images and do not constitute a direct grain size analysis, such as those conducted by Bolliger et al. (2024) on debris-flow deposits. Their method enables the characterization of fine-grained sediments but is limited to a maximum grain size of approximately 16 mm. In contrast, our approach has the potential to extract grain size information for particles larger than 10-15 cm, capturing not only depositional features but also actively moving debris flows. The ability to retrieve GSD data from natural debris flows could be instrumental in improving predictions of future runout and deposition patterns. As highlighted by de Haas et al. (2015), runout and deposition are strongly influenced by flow composition, including water content and the presence of large debris. They further suggest that incorporating GSD into debris flow models could significantly enhance predictive capabilities – an aspect that our method could address effectively.

Lab observations of coarser boulders and rolling boulders traveling faster and catching up with the flow front have not yet been verified. Spielmann et al. (2024) reported that, for the same event, the flow front velocities were 5.5 m/s, 3.4 m/s, and 2.8 m/s at CDG, CD27, and CD29, respectively, with trailing material moving 1.4-1.8× faster than the flow front. At CDG and CD27, we observe velocities exceeding 5 m/s during the first ∼8 minutes of the event, suggesting that debris-flow features may be catching up with the front. Further analysis is required to determine whether this trend is accurate or if larger boulders are being submerged beneath the flow surface. The greater flow depths at CD29 compared to CDG (cf. Fig. 7) may result in the inclusion of objects that could have been detected at CDG or CD27 but are obscured at CD29.

Comparing the object velocities to the flow front velocity at each station reveals a range of velocity ratios (object velocity to front velocity, see Fig. A3), indicating that some objects are traveling faster than the flow front, with ratios between 1.5 and 2.0. Additionally, sorting may be influenced by the quality of detections and sensor configurations. Specifically, the LiDAR sensor at CD27, operating with 64 channels compared to 128 channels at CDG and CD29, results in fewer detections of smaller objects. Understanding these limitations and differences is crucial for refining interpretations of debris-flow dynamics.

6 Conclusions

The new processing method presented in this study successfully implements automatic object detection for the efficient analysis of high-frequency LiDAR point clouds, enabling the calculation of thousands of debris-flow objects, such as boulders, rolling boulders, surge waves, and wood. By leveraging object detection models, the processing time is significantly reduced compared to manual analysis. Our results demonstrate that high-frequency point clouds (10 Hz), combined with a hillshading algorithm, can serve as input for traditional image-based object detection models, achieving high detection accuracies with only 6% of the dataset labeled. Performance metrics across models, including YOLOv5, YOLOv8, YOLOv11, RT-DETR, and Faster R-

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

CNN101, show minimal variation, except for RetinaNet50, which underperforms (e.g., F1 score below 0.8). Notable differences are observed in processing times, with YOLOv11 achieving the highest frame rate at 38.2 FPS, surpassing the 25 FPS threshold for real-time analysis. The use of four distinct object classes – boulders, rolling boulders, surge waves, and wood – proved effective and sufficiently distinct for accurate detection and provided the opportunity to gain deeper insights into flow dynamics. The computed velocities align well with values obtained from alternative methods for the same event, validating the approach. Additionally, the exclusive use of LiDAR data captures a greater number of objects and enables longer tracking durations over more frames, with the added advantage of functioning effectively during nighttime or low-visibility conditions. During an ongoing event, a fining process is observed in the downstream direction, along with variations in particle composition at each monitoring station. For further insights into object velocities and longitudinal sorting, the presented method will be applied to more events.

Code and data availability. Hillshade images, code, training data and model weights will be made available in the ETH Research Collection upon acceptance under a Creative Commons Attribution-ShareAlike 4.0 International license. A DOI for the research collection will be added later.

Video supplement. Video material will be made available in the ETH Research Collection upon acceptance under a Creative Commons Attribution-ShareAlike 4.0 International license. A DOI for the research collection will be added later.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

Appendix A

A1 Evaluation of object detection

Generally, the performance of object detectors is measured by Precision (P) and Recall (R). P describes the accuracy of positive predictions (Equation A1) while R expresses the sensitivity of the model (Equation A2).

$$P = \frac{TP}{TP + FP} \tag{A1}$$

$$R = \frac{TP}{TP + FN} \tag{A2}$$

where TP refers to True Positives, FP to False Positives and FN to False Negatives. To asses both P and R, the F1 score is commonly used. It represents the harmonic mean of P and R, providing a balanced single metric (Equation A3).

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = \frac{2 \cdot P \cdot R}{P + R} \tag{A3}$$

A prediction of the object detection model is considered a TP, if the Intersection over Union (IoU) metric exceeds a defined threshold, 0.5 in our case. The IoU measures the overlap between the predicted and ground truth (GT) bounding box, shown in Equation A4.

$$IoU = \frac{Box_{GT} \cap Box_{Pred}}{Box_{GT} \cup Box_{Pred}} \tag{A4}$$

To evaluate a classifiers performance over all confidence scores, the Average Precision (AP) metrics is used, which is defined the area under Precision-Recall curve. It can be computed by integrating (or approximating the area under) the precision-recall curve with Equation A5:

$$AP = \sum_{n} (R_n - R_{n-1}) \times P_n \tag{A5}$$

where P_n is the precision at the n-th threshold and $R - n$ is the recall at the n-th threshold. The mean Average Precision (mAP) is the mean of the AP scores across all classes in the dataset, providing an overall multi-class measure of model accuracy (Equation A6).

$$mAP = \frac{1}{C} \sum_{c=1}^{C} AP_c \tag{A6}$$

where AP_c is the Average Precision for class c and C is the total number of classes, 4 in our dataset. The mAP can also be evaluated at different IoU thresholds (e.g., mAP@0.5 for IoU of 0.5 or mAP@[0.5:0.95] for an average over multiple IoU values from 0.5 to 0.95 in steps of 0.05).

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

A2 YOLOv11-X Metrics

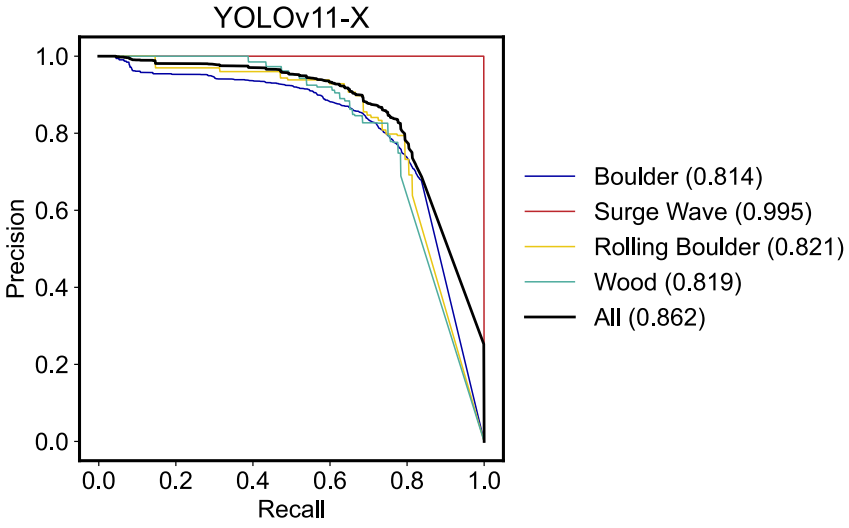

Figure A1. Precision-Recall (PR) curves for the YOLOv11-X performance on the test set across the object classes: Boulder, Surge Wave, Rolling Boulder and Wood (see also Tab. A1). Numbers in parentheses indicate the Average Precision (AP) for each class and the mean Average Precision (mAP) over all classes.

Table A1. Class-wise performance metrics of the YOLOv11-X model evaluated on the test dataset.

Class	Instances	P	P	mAP50	mAP50-95
all	3 476	0.817	0.823	0.862	0.583
Boulder	3 213	0.789	0.755	0.814	0.531
Surge Wave	9	0.9	1	0.995	0.762
Rolling Boulder	102	0.755	0.794	0.821	0.488
Wood	152	0.824	0.741	0.819	0.553

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

A3 Object velocities

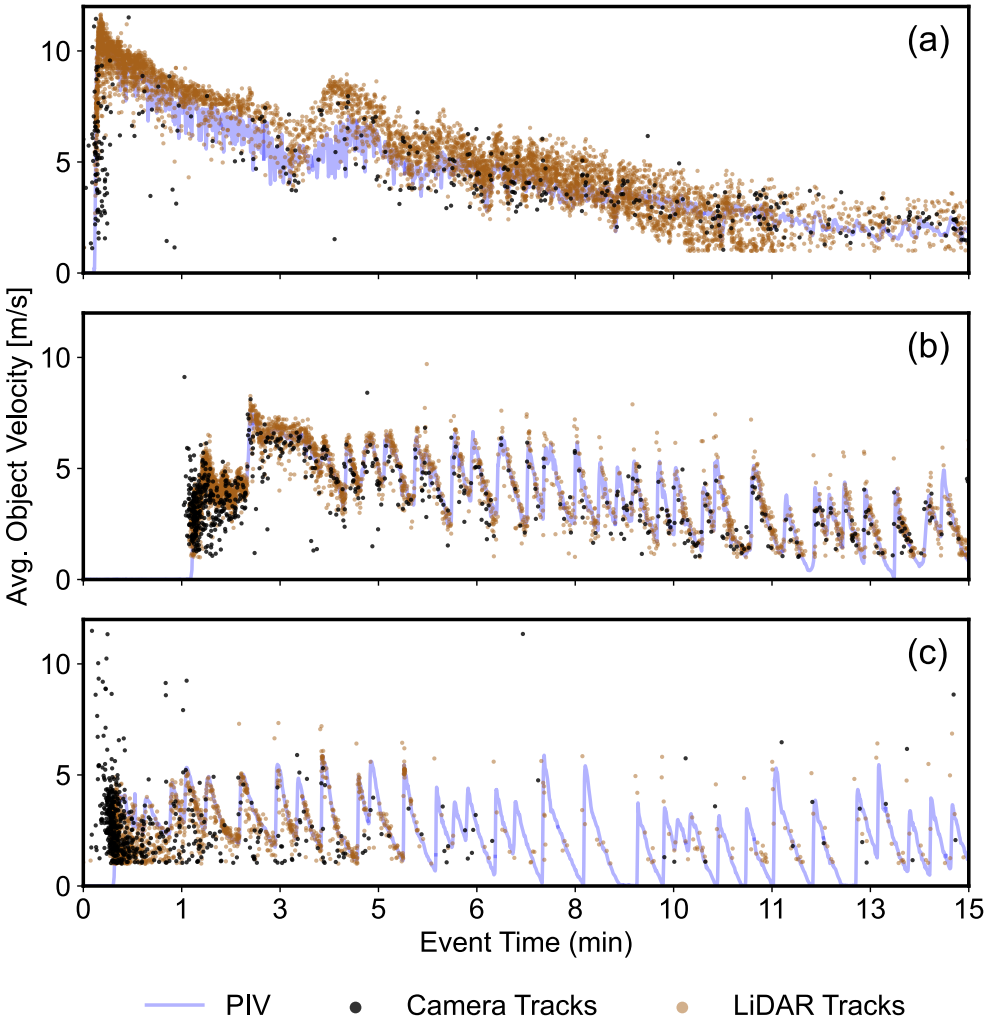

Figure A2. Mean object velocities calculated for the 5 June 2022 event, using the LiDAR hillshade and the camera LiDAR fusion method at the three monitoring stations: (a) Gazoduc (CDG), (b) CD27, and (c) CD29. Surface velocities derived from particle image velocimetry (PIV) by Spielmann et al. (2024) are included as a reference.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

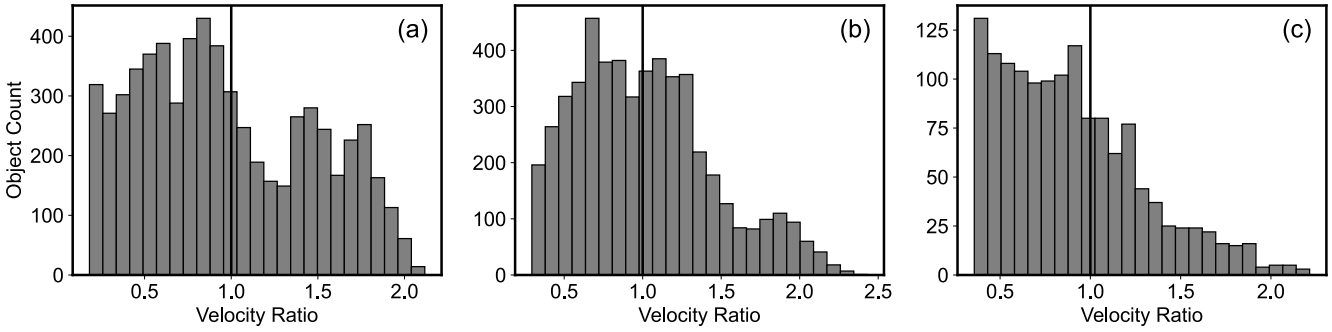

Figure A3. Distribution of object counts as a function of the velocity ratio for (a) CDG, (b) CD27, and (c) CD29. The velocity ratio is defined as $VelocityRatio = ObjectAverageVelocity/FlowFrontVelocity$ for each object at each station. In CDG and CD27, a greater number of objects exhibit velocity ratios exceeding 1 compared to CD29, which also shows overall lower object counts. Note that this figure does not show temporal variations in object velocities relative to the flow front.

A4 Object sizes

Table A2. Particle sizes of the grain size distributions derived from LiDAR detections and object tracks for the 5 June 2022 event for stations CDG, CD27, and CD29. These particle sizes are described using various size metrics, such as D10, which represents the particle size at which 10% of the particles are finer and 90% are coarser.

	Particle Size [m]				
	D10	D30	D50	D60	D90
Detections:					
CDG	0.24	0.3	0.36	0.4	1.18
CD27	0.24	0.3	0.39	0.46	1.15
CD29	0.33	0.42	0.53	0.61	0.84
Object Tracks:					
CDG	0.28	0.34	0.4	0.45	1.25
CD27	0.26	0.32	0.39	0.45	1.07
CD29	0.35	0.41	0.46	0.5	0.76

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

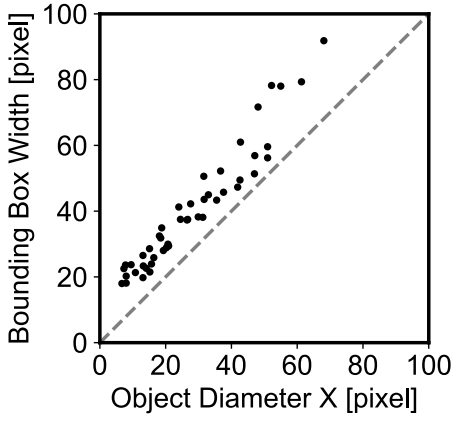

Figure A4. Comparison between object diameter (X-axis) and bounding box width (Y-axis) in pixels, demonstrating a near-linear relationship and highlighting the consistency between manually measured object diameters and bounding box predictions across the dataset. The observed offset of 12.2 pixels, corresponding to 24.4 cm, is likely attributed to the labeling approach, which accounts for turbulence and shadows surrounding the objects.

Author contributions. PES implemented the method, conducted the analysis, produced the figures and wrote the original article draft. PES, JH and JA designed the study. All authors contributed to the interpretation of the results and to the revision of the article.

Competing interests. The contact author has declared that neither they nor their co-authors have any competing interests.

Acknowledgements. The LiDAR installations were funded by the Swiss National Science Foundation (SNSF, grant number 193081) as well as funds from the Chair of Engineering Geology at ETH Zürich. We are grateful for technical support and scientific advice from Stefan Boss (WSL), Christoph Graf (WSL), Dr. Brian McArdell (WSL), Dr. Alexandre Badoux (WSL) and Amanda Åberg (WSL, ETH Zürich).

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

References

Aaron, J., Spielmann, R., McArdell, B. W., and Graf, C.: High-Frequency 3D LiDAR Measurements of a Debris Flow: A Novel Method to Investigate the Dynamics of Full-Scale Events in the Field, Geophysical Research Letters, 50, e2022GL102 373, https://doi.org/10.1029/2022GL102373, 2023.

Åberg, A., Aaron, J., McArdell, B. W., Kirchner, J., de Haas, T., and Hirschberg, J.: Field Validation of the Superelevation Method for Debris-Flow Velocity Estimation Using High-Resolution Lidar and UAV Data, Journal of Geophysical Research: Earth Surface, 129, e2024JF007 857, https://doi.org/10.1029/2024JF007857, 2024.

Aharon, N., Orfaig, R., and Bobrovsky, B.-Z.: BoT-SORT: Robust Associations Multi-Pedestrian Tracking, http://arxiv.org/abs/2206.14651, 2022.

Badoux, A., Graf, C., Rhyner, J., Kuntner, R., and McArdell, B. W.: A Debris-Flow Alarm System for the Alpine Illgraben Catchment: Design and Performance, Natural Hazards, 49, 517–539, https://doi.org/10.1007/s11069-008-9303-x, 2009.

Bennett, G. L., Molnar, P., McArdell, B. W., Schlunegger, F., and Burlando, P.: Patterns and Controls of Sediment Production, Transfer and Yield in the Illgraben, Geomorphology, 188, 68–82, https://doi.org/10.1016/j.geomorph.2012.11.029, 2013.

Berger, C., McArdell, B. W., and Schlunegger, F.: Direct Measurement of Channel Erosion by Debris Flows, Illgraben, Switzerland, Journal of Geophysical Research: Earth Surface, 116, https://doi.org/10.1029/2010JF001722, 2011a.

Berger, C., McArdell, B. W., and Schlunegger, F.: Sediment Transfer Patterns at the Illgraben Catchment, Switzerland: Implications for the Time Scales of Debris Flow Activities, Geomorphology, 125, 421–432, https://doi.org/10.1016/j.geomorph.2010.10.019, 2011b.

Bewley, A.: Abewley/Sort, https://github.com/abewley/sort, 2024.

Bewley, A., Ge, Z., Ott, L., Ramos, F., and Upcroft, B.: Simple Online and Realtime Tracking, in: 2016 IEEE International Conference on Image Processing (ICIP), pp. 3464–3468, IEEE, Phoenix, AZ, USA, ISBN 978-1-4673-9961-6, https://doi.org/10.1109/ICIP.2016.7533003, 2016.

Bolliger, D., Schlunegger, F., and McArdell, B. W.: Comparison of Debris Flow Observations, Including Fine-Sediment Grain Size and Composition and Runout Model Results, at Illgraben, Swiss Alps, Natural Hazards and Earth System Sciences, 24, 1035–1049, https://doi.org/10.5194/nhess-24-1035-2024, 2024.

BSI: Tests for Geometrical Properties of Aggregates. Part 2, Determination of Particle Size Distribution. Test Sieves, Nominal Size of Apertures, British Standards Institution, London, 2020.

de Haas, T., Braat, L., Leuven, J. R. F. W., Lokhorst, I. R., and Kleinhans, M. G.: Effects of Debris Flow Composition on Runout, Depositional Mechanisms, and Deposit Morphology in Laboratory Experiments, Journal of Geophysical Research: Earth Surface, 120, 1949–1972, https://doi.org/10.1002/2015JF003525, 2015.

Diwan, T., Anirudh, G., and Tembhurne, J. V.: Object Detection Using YOLO: Challenges, Architectural Successors, Datasets and Applications, Multimedia Tools and Applications, 82, 9243–9275, https://doi.org/10.1007/s11042-022-13644-y, 2023.

Dowling, C. A. and Santi, P. M.: Debris Flows and Their Toll on Human Life: A Global Analysis of Debris-Flow Fatalities from 1950 to 2011, Natural Hazards, 71, 203–227, https://doi.org/10.1007/s11069-013-0907-4, 2014.

Edwards, A. N. and Gray, J. M. N. T.: Erosion–Deposition Waves in Shallow Granular Free-Surface Flows, Journal of Fluid Mechanics, 762, 35–67, https://doi.org/10.1017/jfm.2014.643, 2015.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

Fernandes, D., Silva, A., Névoa, R., Simões, C., Gonzalez, D., Guevara, M., Novais, P., Monteiro, J., and Melo-Pinto, P.: Point-Cloud Based 3D Object Detection and Classification Methods for Self-Driving Applications: A Survey and Taxonomy, Information Fusion, 68, 161–191, https://doi.org/10.1016/j.inffus.2020.11.002, 2021.

Gray, J. M. N. T.: Particle Segregation in Dense Granular Flows, Annual Review of Fluid Mechanics, 50, 407–433, https://doi.org/10.1146/annurev-fluid-122316-045201, 2018.

Hilker, N., Badoux, A., and Hegg, C.: The Swiss Flood and Landslide Damage Database 1972–2007, Natural Hazards and Earth System Sciences, 9, 913–925, https://doi.org/10.5194/nhess-9-913-2009, 2009.

Hirschberg, J. and Aaron, J.: Deep-Learning-Based 3D Surge-Wave Detection and Tracking in Debris Flows Using Cameras and LiDARs, in: 14th International Symposium on Landslides (ISL 2024), https://www.research-collection.ethz.ch/handle/20.500.11850/718419, 2024.

Hirschberg, J., Badoux, A., McArdell, B. W., Leonarduzzi, E., and Molnar, P.: Evaluating Methods for Debris-Flow Prediction Based on Rainfall in an Alpine Catchment, Natural Hazards and Earth System Sciences, 21, 2773–2789, https://doi.org/10.5194/nhess-21-2773-2021, 2021a.

Hirschberg, J., Fatichi, S., Bennett, G. L., McArdell, B. W., Peleg, N., Lane, S. N., Schlunegger, F., and Molnar, P.: Climate Change Impacts on Sediment Yield and Debris-Flow Activity in an Alpine Catchment, Journal of Geophysical Research: Earth Surface, 126, e2020JF005 739, https://doi.org/10.1029/2020JF005739, 2021b.

HumanSignal: Open Source Data Labeling, https://labelstud.io/, 2024.

Hungr, O., Leroueil, S., and Picarelli, L.: The Varnes Classification of Landslide Types, an Update, Landslides, 11, 167–194, https://doi.org/10.1007/s10346-013-0436-y, 2014.

Hürlimann, M., Rickenmann, D., and Graf, C.: Field and Monitoring Data of Debris-Flow Events in the Swiss Alps, Canadian Geotechnical Journal, 40, 161–175, https://doi.org/10.1139/t02-087, 2003.

Hürlimann, M., Coviello, V., Bel, C., Guo, X., Berti, M., Graf, C., Hübl, J., Miyata, S., Smith, J. B., and Yin, H.-Y.: Debris-Flow Monitoring and Warning: Review and Examples, Earth-Science Reviews, 199, https://doi.org/10.1016/j.earscirev.2019.102981, 2019.

Iverson, R. M.: The Physics of Debris Flows, Reviews of Geophysics, 35, 245–296, https://doi.org/10.1029/97RG00426, 1997.

Iverson, R. M., Logan, M., LaHusen, R. G., and Berti, M.: The Perfect Debris Flow? Aggregated Results from 28 Large-Scale Experiments, Journal of Geophysical Research F: Earth Surface, 115, F03 005, https://doi.org/10.1029/2009JF001514, 2010.

Jacquemart, M., Weber, S., Chiarle, M., Chmiel, M., Cicoira, A., Corona, C., Eckert, N., Gaume, J., Giacona, F., Hirschberg, J., Kaitna, R., Magnin, F., Mayer, S., Moos, C., van Herwijnen, A., and Stoffel, M.: Detecting the Impact of Climate Change on Alpine Mass Movements in Observational Records from the European Alps, Earth-Science Reviews, 258, 104 886, https://doi.org/10.1016/j.earscirev.2024.104886, 2024.

Jakob, M. and Hungr, O.: Introduction, in: Debris-Flow Hazards and Related Phenomena, pp. 1–7, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-540-20726-9 978-3-540-27129-1, https://doi.org/10.1007/3-540-27129-5_1, 2005.

Jocher, G.: YOLOv5 by Ultralytics, https://github.com/ultralytics/yolov5, 2020.

Jocher, G. and Qiu, J.: YOLO11 by Ultralytics, https://github.com/ultralytics/ultralytics, 2024.

Jocher, G., Chaurasia, A., and Qiu, J.: YOLOv8 by Ultralytics, https://github.com/ultralytics/ultralytics, 2023.

Kaitna, R., Prenner, D., Switanek, M., Maraun, D., Stoffel, M., and Hrachowitz, M.: Changes of Hydro-Meteorological Trigger Conditions for Debris Flows in a Future Alpine Climate, Science of The Total Environment, 872, 162 227, https://doi.org/10.1016/j.scitotenv.2023.162227, 2023.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

Kean, J. W., McCoy, S. W., Tucker, G. E., Staley, D. M., and Coe, J. A.: Runoff-Generated Debris Flows: Observations and Modeling of Surge Initiation, Magnitude, and Frequency, Journal of Geophysical Research: Earth Surface, 118, 2190–2207, https://doi.org/10.1002/jgrf.20148, 2013.

Kokalj, Ž. and Somrak, M.: Why Not a Single Image? Combining Visualizations to Facilitate Fieldwork and On-Screen Mapping, Remote Sensing, 11, 747, https://doi.org/10.3390/rs11070747, 2019.

Kokalj, Ž., Maroh, Ž., Oštir, K., Zakšek, K., and Čož, N.: EarthObservation/RVT_py: Relief Visualization Toolbox in Python, https://github.com/EarthObservation/RVT_py, 2022.

Li, Y., Li, S., Du, H., Chen, L., Zhang, D., and Li, Y.: YOLO-ACN: Focusing on Small Target and Occluded Object Detection, IEEE Access, 8, 227 288–227 303, https://doi.org/10.1109/ACCESS.2020.3046515, 2020.

Lichtenhahn, C.: Zwei Betonmauern: Die Geschieberückhaltesperre Am Illgraben [Wallis] Und Die Staumauer Des Hochwasserschutzbeckens an Der Orlegna Im Bergell [Graubünden], in: Internationales Symposium Interpraevent, vol. F.f.v. Hochwasserbekämpfung, pp. 451–456, Villach, Austria, 1971.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P.: Focal Loss for Dense Object Detection, https://doi.org/10.48550/arXiv.1708.02002, 2018.

Liu, Y., Sun, P., Wergeles, N., and Shang, Y.: A Survey and Performance Evaluation of Deep Learning Methods for Small Object Detection, Expert Systems with Applications, 172, 114 602, https://doi.org/10.1016/j.eswa.2021.114602, 2021.

McArdell, B. W.: Field Measurements of Forces in Debris Flows at the Illgraben: Implications for Channel-Bed Erosion, International Journal of Erosion Control Engineering, 9, 194–198, https://doi.org/10.13101/ijece.9.194, 2016.

McArdell, B. W. and Sartori, M.: The Illgraben Torrent System, in: Landscapes and Landforms of Switzerland, edited by Reynard, E., pp. 367–378, Springer International Publishing, Cham, ISBN 978-3-030-43201-0 978-3-030-43203-4, https://doi.org/10.1007/978-3-030-43203-4_25, 2021.

McArdell, B. W., Bartelt, P., and Kowalski, J.: Field Observations of Basal Forces and Fluid Pore Pressure in a Debris Flow, Geophysical Research Letters, 34, https://doi.org/10.1029/2006GL029183, 2007.

Pouliquen, O., Delour, J., and Savage, S. B.: Fingering in Granular Flows, Nature, 386, 816–817, https://doi.org/10.1038/386816a0, 1997.

Prakash, N., Santi, P., Strouth, A., Sepulveda, S. A., and Dowling, C.: Fatalities from Debris Flows: Worldwide Distribution and Trends, in: Advances in Debris-flow Science and Practice, edited by Jakob, M., McDougall, S., and Santi, P., pp. 75–91, Springer International Publishing, Cham, ISBN 978-3-031-48690-6 978-3-031-48691-3, https://doi.org/10.1007/978-3-031-48691-3_3, 2024.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J.: PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, http://arxiv.org/abs/1612.00593, 2017.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection, https://doi.org/10.48550/arXiv.1506.02640, 2016.

Ren, S., He, K., Girshick, R., and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, https://doi.org/10.48550/arXiv.1506.01497, 2016.

Ruis, F. A., Liezenga, A. M., Heslinga, F. G., Ballan, L., Eker, T. A., den Hollander, R. J. M., van Leeuwen, M. C., Dijk, J., and Huizinga, W.: Improving Object Detector Training on Synthetic Data by Starting With a Strong Baseline Methodology, http://arxiv.org/abs/2405.19822, 2024.

Shorten, C. and Khoshgoftaar, T. M.: A Survey on Image Data Augmentation for Deep Learning, Journal of Big Data, 6, 60, https://doi.org/10.1186/s40537-019-0197-0, 2019.

https://doi.org/10.5194/egusphere-2025-743
Preprint. Discussion started: 28 May 2025

Spielmann, R. and Aaron, J.: A New Method for Detailed Discharge and Volume Measurements of Debris Flows Based on High-Frequency 3D LiDAR Point Clouds; Illgraben, Switzerland, Engineering Geology, 329, 107 386, https://doi.org/10.1016/j.enggeo.2023.107386, 2024.

Spielmann, R., Huber, S., and Aaron, J.: Direct Measurements of Debris-Flow Feature Velocities Using High-Frequency 3D LiDAR Scanners, in: XIVth International Symposium on Landslides, 2024.

Terven, J., Córdova-Esparza, D.-M., and Romero-González, J.-A.: A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS, Machine Learning and Knowledge Extraction, 5, 1680–1716, https://doi.org/10.3390/make5040083, 2023.

Thielicke, W. and Sonntag, R.: Particle Image Velocimetry for MATLAB: Accuracy and Enhanced Algorithms in PIVlab, Journal of Open Research Software, 9, 12, https://doi.org/10.5334/jors.334, 2021.

Thornton, A. R. and Gray, J. M. N. T.: Breaking Size Segregation Waves and Particle Recirculation in Granular Avalanches, Journal of Fluid Mechanics, 596, 261–284, https://doi.org/10.1017/S0022112007009445, 2008.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girshick, R.: Detectron2, https://github.com/facebookresearch/detectron2, 2019.

Yang, B., Luo, W., and Urtasun, R.: PIXOR: Real-time 3D Object Detection from Point Clouds, https://doi.org/10.48550/ARXIV.1902.06326, 2019.

Zanuttigh, B. and Lamberti, A.: Instability and Surge Development in Debris Flows, Rev. Geophys, 45, https://doi.org/10.1029/2005RG000175, 2007.

Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., and Chen, J.: DETRs Beat YOLOs on Real-time Object Detection, http://arxiv.org/abs/2304.08069, 2024.

Zhou, Q.-Y., Park, J., and Koltun, V.: Open3D: A Modern Library for 3D Data Processing, http://arxiv.org/abs/1801.09847, 2018.

Zou, Z., Chen, K., Shi, Z., Guo, Y., and Ye, J.: Object Detection in 20 Years: A Survey, Proceedings of the IEEE, 111, 257–276, https://doi.org/10.1109/JPROC.2023.3238524, 2023.