https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

Glacier surface velocity from supraglacial boulder tracking: can a neural network replace manual tracking on debris-covered glaciers?

Francesco Parizia[1,2,*], **Alessandro Farbo**[1,*], **Marco Giardino**[3], **Luigi Perotti**[1]

[*] These authors contributed equally to this work and share first authorship.

[1] Dipartimento di Scienze Agrarie, Forestali e Alimentari (DISAFA), Università degli Studi di Torino, Largo Paolo Braccini 2, 10095 Grugliasco (TO), Italy
[2] Dipartimento di Ingegneria Civile, Edile e Ambientale (DICEA), Sapienza Università di Roma, Via Eudossiana 18, 00184 Roma, Italy
[3] Dipartimento di Scienze della Terra (DST), Università degli Studi di Torino, Via Valperga Caluso 35, 10125 Torino, Italy

Correspondence: Francesco Parizia (francesco.parizia@unito.it), Alessandro Farbo (alessandro.farbo@unito.it)

Keywords: glacier surface velocity; feature tracking; deep learning; orthophotos;

Abstract

Tracking supraglacial boulders on repeat imagery is a long-established, but still manual, way to measure surface velocity on debris-covered glaciers. As Alpine glaciers thin and slow down, knowing whether a tongue still flows is increasingly relevant. As a first step toward automating this procedure, we present a neural network that relocates an operator-selected boulder in a later orthophoto. Boulders are delineated with SAM2 and matched by a ConvNeXt network trained on manual tracks from Miage and Ciardoney (Italy) and Gorner or Zmutt (Switzerland). Predictions can be reranked with a coarse DEM and refined with a consensus among neighboring boulders (local context). We evaluated the network with two leave-one-glacier-out experiments and on Belvedere Glacier (Italy), never used for training or model selection. Over 2020–2023 on Belvedere, terrain reranking with local context relocated 39 and 35 of 55 boulders (71 % and 64 % per fold) within 2 m. Single-query predictions (56–58 %) matched the best of four frozen benchmarks, normalized cross-correlation (58 %). Over 5 and 8 years all methods failed for most boulders, but the network relocated 13–14 of 47 and 8–10 of 46, against at most 7 and 1 for the frozen benchmarks. Correct matches reproduced manual surface speeds within 0.5 m yr^{-1}, and with local context no boulder was assigned to the opposite flow regime (fast ≥ 17, slow < 7 m yr−1) over 2020–2023. The network does not yet replace the operator: it needs a starting point per boulder, cannot flag its own wrong matches and was tested on a single unseen glacier.

1 Introduction

Rock debris covers about 7.3 % of the world's mountain glacier area and is present on 44 % of Earth's glaciers (Herreid and Pellicciotti, 2020). On these debris-covered glaciers, surface velocity is a primary indicator of dynamic state. Supraglacial boulders move with the ice and are the most direct tracers available.

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

Tracking supraglacial boulders to measure surface velocity is a long-standing practice in Alpine glaciology. On Miage Glacier, Lesca (1974a) used analytical photogrammetry to measure surface velocity by tracking a large number of boulders on the ablation tongue. He noted that the method suits valley glaciers with surface slopes of 20 % or less and a supraglacial ("floating") moraine. He checked the photogrammetric velocities against precise theodolite and electronic distance measurements (Lesca, 1974b), and derived ice thickness from them with Somigliana's formula, in agreement with seismic soundings (Lesca, 1974a). The resulting maps of displacement and annual velocity on Miage cover 1965–1970, 1970–1971 and 1965–1971 at 1:10 000. The same approach, applied to boulders on repeat aerial imagery, later reconstructed more than five decades (1961–2017) of flow velocity on the tongue of Zmuttgletscher (Mölg et al., 2019). On Belvedere Glacier, about 130 rock features (115 in 2020) were collimated manually on UAV orthophotos to derive annual velocities, with an uncertainty of about 0.3 m yr^{-1} (Ioli et al., 2022). Manual boulder tracking on repeat orthophotos is still used in recent work: on the debris-covered Nautárdalsjökull (northern Iceland), boulder centroids were digitized on orthophotos of 1994, 2009 and 2021 to derive surface velocities (Tanarro et al., 2026). On the Plator rock glacier (Italian Alps), large displacements and the lack of sharpness of the 1981–1999 regional orthophotos prevented normalized cross-correlation, so boulders were tracked manually, recognized by their shape, size and position relative to each other and to the main landforms; automatic tracking became possible only from 2003 onward, and only in the slower parts of the landform (Scotti et al., 2017). Fifty years on, the tracers are the same and the identification is still largely done by hand.

Flow is also part of what defines a glacier. The widely used definition of Cogley et al. (2011) requires ice "showing evidence of past or present flow". As glaciers vanish, whether ice still flows has become one of several practical criteria, together with area, thickness and crevassing, used to decide whether a remnant is still a glacier, although the standard definition also accepts past flow (Pope, 2025). Many Alpine debris-covered tongues are slowing down. On Miage Glacier, surface velocity decreased by 46 % between 1990 and 2018 while the glacier kept losing mass (Stefaniak et al., 2021). Since 2005, the lowermost 1.5 km of Zmuttgletscher has been quasi-stagnant, with velocities below 3 m yr^{-1} (Mölg et al., 2019). Detecting the transition from active flow to stagnation requires repeated, spatially dense displacement measurements.

Debris supply to glacier surfaces is changing as well. Debris cover on Zmuttgletscher grew from about 13 % to 32 % of the glacier area between 1859 and 2013 (Mölg et al., 2019). On Miage, debris cover expanded by about 8.5 % of glacier area over 1990–2018 (Stefaniak et al., 2021). Rockwalls freshly exposed by glacier recession are a major rockfall source. At the Kitzsteinhorn, about 60 % of the rockfall volume recorded in 2011–2017 detached from less than 10 vertical metres above the glacier surface (Hartmeyer et al., 2020). More boulders on glacier surfaces means more potential tracers, but also more surface change between surveys.

High-resolution imagery is becoming easier to obtain. In Switzerland, the SWISSIMAGE mosaic is updated on a 3-year cycle at 10 cm resolution in the main valleys and 25 cm over the Alps, and has been open data since 2021, with acquisitions from 2017 onward available for download (swisstopo, 2026b). UAV surveys now routinely produce 10 cm orthophotos of entire glaciers or glacier tongues (Benoit et al., 2019; Ioli et al., 2022). The archive of repeat orthophotos over debris-covered glaciers is therefore growing faster than it can be analyzed by hand.

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

Area-based image matching (normalized cross-correlation, phase correlation, COSI-Corr) is the standard way to derive glacier displacements from optical imagery (Heid and Kääb, 2012). On debris-covered surfaces, however, low contrast and surface heterogeneity remain persistent problems (Norova et al., 2026). Learned feature matchers from computer vision (SuperPoint with SuperGlue or LightGlue) have recently outperformed cross-correlation on UAV imagery of Austerdalsbreen, a mostly bare-ice outlet glacier (Zandler et al., 2025). These methods return dense or keypoint-based displacement fields. They do not answer the question that manual boulder tracking answers: where did this specific boulder go?

Object-level tracking of boulders has so far been developed mainly for rock glaciers and landslides, and mostly on 3D data. On the Äußeres Hochebenkar rock glacier, corresponding planar boulder faces are matched between terrestrial laser-scanning point clouds (Hartl et al., 2023), and boulders segmented in UAV-LiDAR point clouds have been tracked with a neighbourhood-based matching approach (Carniel et al., 2025). On monoscopic time-lapse imagery, the deep-learning point tracker PIPs++ has followed surface features, including boulders, on a fast-moving rock glacier and a landslide (Hendrickx et al., 2025). In orthophoto-based studies, boulders typically serve as GNSS-surveyed markers to validate area-based matching (e.g. Vivero et al., 2022). On debris-covered glaciers, deep learning has so far been applied mainly to mapping the extent of debris cover (Kaushik et al., 2022, 2025), not to measuring motion. Foundation models such as SAM segment glaciological features without task-specific training, although their performance depends on prompting (Shankar et al., 2024), but we found no study that uses them to follow individual supraglacial boulders. To our knowledge, automated tracking of individual boulders on debris-covered glaciers from repeat orthophotos has not yet been attempted.

Here we ask whether automated tracking can relocate individual boulders on an unseen debris-covered glacier as reliably as manual tracking and established image-matching methods. We train a correspondence network on manually tracked boulders from four Alpine glaciers, using orthophotos and, optionally, a coarse global DEM, and combine its predictions with terrain reranking and a consensus among neighboring boulders. We then test it on held-out glaciers and on Belvedere Glacier, never seen in training, over survey intervals of three to eight years, and compare it with four established matching methods.

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

2 Methods

2.1 Study sites, imagery and reference data

We use five glaciers in the western Alps: four for training and cross-validation, and Belvedere as an independent transfer site (Table 1, Fig. 1).

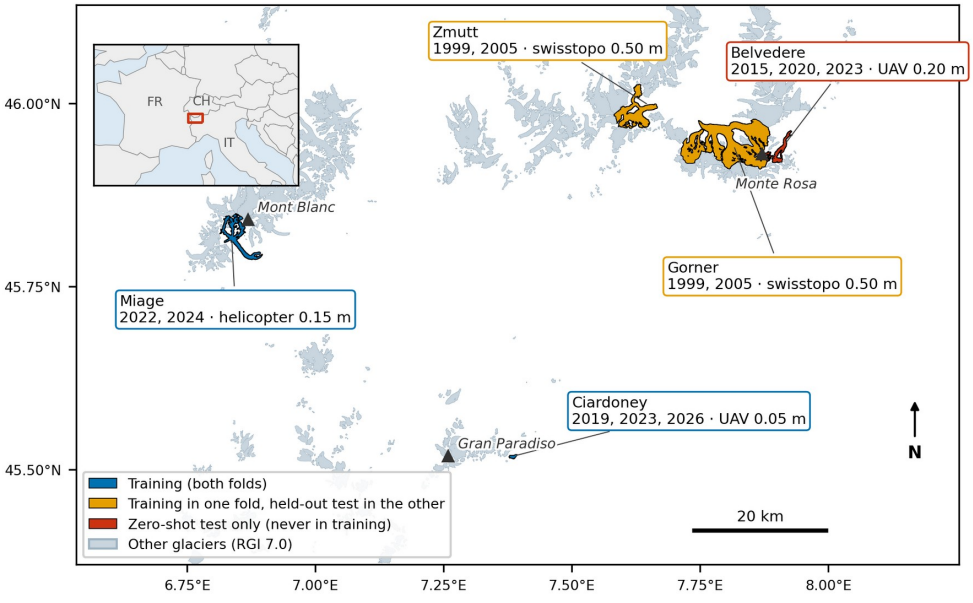


Figure 1. *Study glaciers in the western Alps and their role in the two leave-one-glacier-out experiments: Miage and Ciardoney (blue) are used for training in both folds; Gorner and Zmutt (orange) are used for training in one fold and as held-out validation and test glacier in the other; Belvedere (red) is used only for zero-shot testing. Labels give the survey years and the imagery source and ground sampling distance. For Gorner and Zmutt the outline covers the whole glacier, but only the debris-covered terminal tongue, with supraglacial ("floating") moraines, was used for tracking. Glacier outlines from RGI 7.0 (RGI Consortium, 2023); country boundaries in the inset from Natural Earth (version 5.1.1).*

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 1. *Glaciers, surveys and imagery, and role of each glacier in the experiments.*

Glacier	Country	Role	Surveys (years)	Imagery source, GSD
Miage	Italy	Training (both folds)	2022, 2024	Helicopter photogrammetry, 0.15 m
Ciardoney	Italy	Training (both folds)	2019, 2023, 2026	UAV orthophotos, 0.05 m
Gorner	Switzerland	Training (fold 1); validation + test (fold 0)	1999, 2005	swisstopo color aerial images (LUBIS; swisstopo, 2026a), orthophotos by swisstopo, 0.50 m
Zmutt	Switzerland	Training (fold 0); validation + test (fold 1)	1999, 2005	swisstopo color aerial images (LUBIS), orthophotos by swisstopo, 0.50 m
Belvedere	Italy	Zero-shot test only	2015, 2020, 2023	UAV orthophotos from the open dataset of Ioli et al. (2024), 0.20 m

Reference displacements were obtained by manual tracking, the procedure that the network is meant to reproduce. On each orthophoto, the centroid of every selected boulder was digitized in QGIS by visual interpretation, and the centroids of the same boulder in successive surveys were linked into a track. On Gorner and Zmutt, tracking was restricted to the debris-covered terminal tongues, where supraglacial ("floating") moraines provide abundant boulders; the upper, largely debris-free parts of these glaciers were not used. On Ciardoney, 91 centroids were digitized across the 2019, 2023 and 2026 surveys. In total, the training glaciers provide nine surveys, six chronological survey pairs and 603 boulder observations (Table 1). The digitization uncertainty has not been quantified yet.

On Belvedere, 55 boulders were tracked between 2020 and 2023, but only 47 and 46 in the pairs that include 2015. The difference is mainly due to fresh snow in the October 2015 orthophoto, which covered part of the tongue and made some boulders impossible to digitize (Fig. 2). The three Belvedere query sets therefore differ slightly, and success rates across pairs are compared as proportions of their own query set.

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

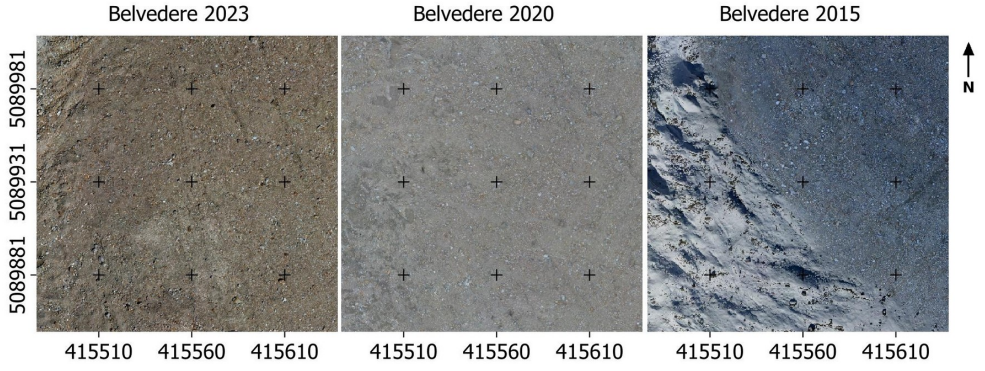


Figure 2. *The same 150 m × 150 m sector of the Belvedere tongue in the 2023, 2020 and 2015 UAV orthophotos (Ioli et al., 2024). The 2023 image (24–26 July) is sharp and homogeneous; the 2020 image (26–27 July) is homogeneous in illumination but locally smeared; the 2015 image (8 and 23 October) combines a different color balance, long shadows and fresh snow that hides boulders in the left half of the sector. Coordinates are in EPSG 32632.*

Terrain information was taken mainly from the Copernicus DEM GLO-30 (ESA, 2021), a global digital surface model with a posting of 1 arcsec (about 30 m) derived from TanDEM-X acquisitions of 2011–2015 and with a specified absolute vertical accuracy better than 4 m (LE90); the SRTMGL1 DEM (NASA JPL, 2013), acquired in February 2000, was used only as an auxiliary product, to check the downslope direction derived from Copernicus; it supplies no elevations and the two DEMs are not merged (details in Sect. 2.4). The Copernicus DEM is not contemporaneous with the orthophotos: it post-dates the Gorner and Zmutt surveys (1999–2005) and pre-dates the other surveys (2015–2026) by up to about 15 years, so its surface does not reproduce the glacier thinning in between. At 30 m it also does not resolve the hummocky relief of the debris cover. It was therefore used only as a regional prior on slope and down-glacier direction (Sect. 2.4), and never to constrain endpoints at the scale of individual boulders.

2.2 Tracking model

The operator supplies the center of a boulder in the earlier (source) orthophoto, and the network searches the later (target) orthophoto for the same boulder, returning the position with the highest matching score (Fig. 3). As in manual tracking, the network does not detect boulders by itself, and it always returns a location.

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

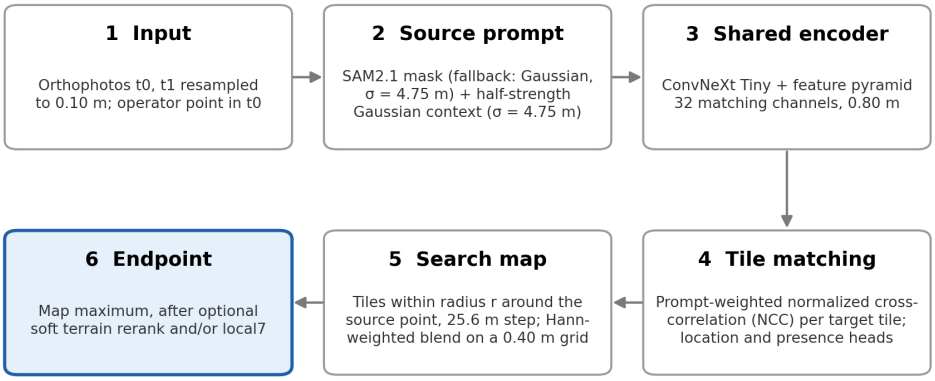


Figure 3. *Inference for a single query boulder. The operator supplies the boulder center in the earlier orthophoto; the network matches the prompt-weighted source template against overlapping target tiles within a search radius r around the source point, blends the tile scores into one search map and returns its maximum, optionally after soft terrain reranking and/or the local7 consensus (Sect. 2.4).*

The network works in three steps (architecture in Fig. S1). First, the source boulder is outlined with the pretrained segmentation model SAM2.1 (Ravi et al., 2024), prompted with the supplied point; the mask tells the network which part of the source image belongs to the boulder, while the surrounding texture remains visible. Second, the source and target images are converted into feature maps by the same convolutional encoder, ConvNeXt Tiny pretrained on natural images (Liu et al., 2022), followed by a feature pyramid that produces features every 0.80 m. Third, the source features, weighted by the mask, are compared with every position of the target by mask-weighted normalized cross-correlation, and two small convolutional layers turn this similarity into a score map. The maximum of the score map is the predicted endpoint.

The search area is a circle of fixed radius around the source point (Sect. 2.5). It is covered by overlapping target tiles of 51.2 m × 51.2 m, whose score maps are blended into a single map on a 0.40 m grid (Fig. 3). Patch sizes, preprocessing, tiling and the other settings are given in Sect. S2.1 and Table S1.

2.3 Training

The network is trained on the manual tracks of the training glaciers (Table 1) in two stages. In the first stage, it learns to locate the boulder in single target tiles that either contain the manual endpoint or exclude it by at least 2 m (2,576 tile pairs). In the second stage, the encoder and feature pyramid are kept fixed and the matching layers are trained on complete searches within a 160 m radius, so that the true endpoint ranks above every other boulder in the search area. In both stages, the training objective rewards a sharp score peak at the manual endpoint and penalizes competing peaks elsewhere. To make the network robust to changes between surveys, training images are rotated and flipped, and their brightness, color, sharpness, noise and shadows are altered independently for the two dates; coarser imagery is also simulated by downsampling. After each training epoch, the model is evaluated on the validation part of the held-out glacier (Sect. 2.5), and the version that performs best on these data,

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

including the starting model before that stage, is retained; test data are never used for selection. The loss functions are given in Sect. S2.2 and all hyperparameters in Table S1.

2.4 Terrain information and local context

Boulders generally move down-glacier, so we also tested whether a coarse terrain prior helps to discard implausible matches. The prior is derived from the Copernicus DEM GLO-30 (ESA, 2021), smoothed over a few hundred meters, which gives the general downslope direction and the elevation. A second downslope direction, from the SRTMGL1 DEM (NASA JPL, 2013), is used only to weaken the prior where the two DEMs disagree; the prior is also weakened on flat terrain. Because the DEMs are coarse and not contemporaneous with the imagery (Sect. 2.1), the prior acts as a soft preference: displacements pointing up-glacier or uphill are penalized but not excluded, since boulders can slide off ice cliffs and hummocks.

We compared three ways of adding the prior to the image-only network. Soft rerank subtracts an up-glacier and uphill penalty from the score map at inference, with weights chosen on validation data and no retraining. Physics loss adds a similar penalty to the training objective during a short fine-tuning; a matched control repeats the fine-tuning without it, so that the effect of the penalty can be separated from that of the extra training. Terrain head is a small network that learns a correction of the score map from the terrain variables. The formulation of the three variants is given in Sect. S2.3.

Finally, we tested a local-context step, called local7, which imitates the way operators use neighboring boulders. For each boulder, the median displacement predicted for its seven nearest neighbors in the same survey pair gives a local estimate, and scores far from the corresponding endpoint are lowered before the maximum is taken again (details in Sect. S2.4). Local7 uses no manual endpoints, but each result depends on the other boulders tracked in the same pair, so local7 results are reported separately from single-query results.

2.5 Baselines and evaluation

We compare the network with four established image-matching methods, used without retraining: grayscale normalized cross-correlation (NCC), gradient NCC, SIFT (Lowe, 2004) and DISK (Tyszkiewicz et al., 2020) with LightGlue (Lindenberger et al., 2023). They run on a 0.20 m grid and serve as practical references, not as a controlled comparison of architectures.

We run two leave-one-glacier-out experiments. Fold 0 trains on Miage, Ciardoney and Zmutt and holds out Gorner (13 validation and 13 test boulders); fold 1 trains on Miage, Ciardoney and Gorner and holds out Zmutt (22 validation and 23 test boulders). The held-out boulders are split spatially into validation and test halves, and validation is used only to select models and terrain parameters. Miage and Ciardoney stay in training in both folds: Miage provides most of the tracks (184) over a wide range of displacements, and Ciardoney is the only training glacier with three surveys, spanning short and long intervals. Gorner and Zmutt, imaged with the same swisstopo product but with test boulders moving at about 14 and 4 m yr^{-1} (Sect. 4.3), are alternated, so that each fold tests transfer to a glacier with different dynamics.

Belvedere is used only for zero-shot testing: its labels were never used for training or model selection and were read only after the predictions had been saved. The search radii, 75 m for 2020→2023, 125 m for 2015→2020 and 200 m for 2015→2023, were set from the training displacements and from the

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

highest surface velocity published for Belvedere, about 22 m yr^{-1} (Ioli et al., 2022), plus a margin of about 13 %; a posteriori, every manual endpoint lay within its search radius. The two fold models are applied separately, and each pair is predicted directly, without chaining 2015→2023 through 2020.

The main metric is endpoint success, the share of queries (boulders tracked in a survey pair) whose predicted endpoint lies within 2 m of the manual one, with Wilson 95 % intervals; we also report median and 90th-percentile errors. For benchmarks that can return no match, missing endpoints count as failures. As a plausibility check, we count predictions pointing up-glacier or ending more than 5 m above the source. Surface speed is the displacement divided by the time between mid-acquisition dates (mid-October 2015, late July 2020 and 2023), and boulders are classed as slow (< 7 m yr^{-1}), intermediate (7–17 m yr^{-1}) or fast (≥ 17 m yr^{-1}), following the two regimes reported by Ioli et al. (2022).

3 Results

3.1 Held-out glaciers

On held-out glaciers the image-only network relocated 4 of 13 Gorner boulders (Wilson 95 % interval 13–58 %) and 10 of 23 Zmutt boulders (26–63 %) within 2 m (Figs. 4 and S2). On Zmutt, soft reranking raised success to 15/23 and cut the median error from 81.2 m to 0.79 m. On Gorner it left success unchanged (4/13) and lowered the mean error from 94.6 m to 85.0 m. Local context gave the best results on both glaciers: 8 of 13 on Gorner with the image-only model (Wilson 95 % interval 36–82 %) and 18 of 23 on Zmutt with the soft rerank (58–90 %). On Zmutt the frozen baselines also performed well: gradient NCC reached 18/23, the best single-query result on that glacier and equal to soft rerank + local7, whereas on Gorner no frozen benchmark relocated more than 2 of 13 boulders.

The physics-loss model and its zero-physics control returned identical endpoints for all 36 held-out queries. The terrain head moved 12 of 13 Gorner endpoints but none of the 23 Zmutt endpoints. In fold 0, validation selected the starting model (epoch 0) for both runs, so both reproduce the search model; in fold 1, both runs selected epoch 3, and the physical term did not change any held-out endpoint. Likewise, the terrain head of fold 1 selected its zero-correction starting model, which is identical to the image-only model, whereas that of fold 0 selected epoch 1.

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

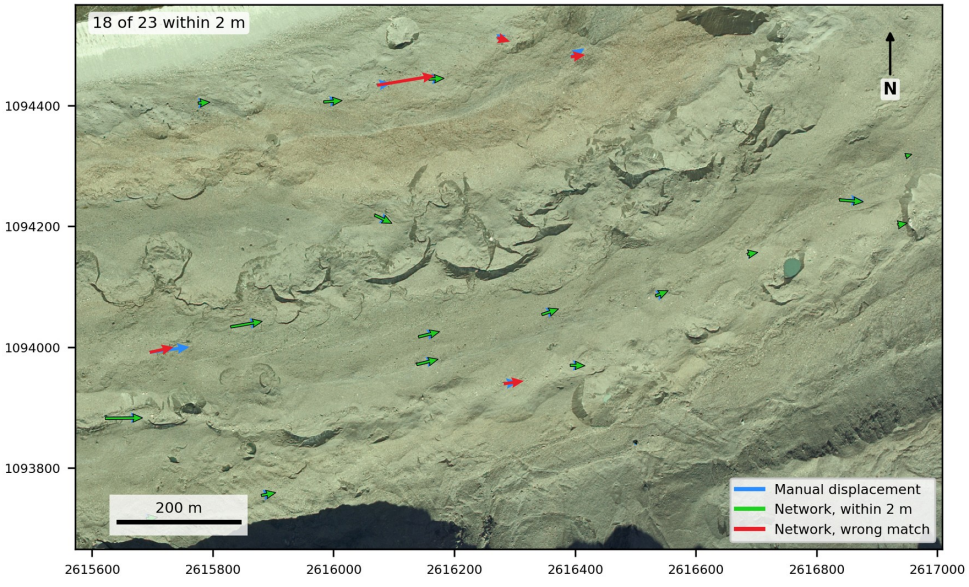


Figure 4. *Zmutt held-out test (1999→2005, swisstopo orthophotos; background: 1999 orthophoto), fold 1 model (trained without Zmutt) with soft terrain rerank and local7 context: manual displacement vectors (blue) and predicted vectors within 2 m of the manual endpoint (green) or wrong matches (red); 18 of 23 endpoints within 2 m. Coordinates are in EPSG 2056. The held-out test of Gorner is shown in Fig. S2.*

3.2 Transfer to Belvedere

On Belvedere, which was never used for training or model selection, the best result of the main pipeline was obtained by combining soft terrain reranking with local context: 39 of 55 boulders (71 %, model trained without Zmutt; Wilson 95 % interval 58–81 %) and 35 of 55 (64 %, model trained without Gorner; 50–75 %) were relocated within 2 m over 2020→2023 (Figs. 5 and 6; maps for all pairs and folds in Fig. S3). Success decreased with the time between surveys, although in this dataset the interval covaries with image quality (the 2015 survey) and with the search radius. For the 3-year pair, the image-only models placed 30/55 endpoints within 2 m in both folds, with median errors of 0.62 m (Gorner fold) and 1.12 m (Zmutt fold). For the 8-year pair 2015→2023, they placed 7/46 and 3/46, with median errors above 130 m.

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

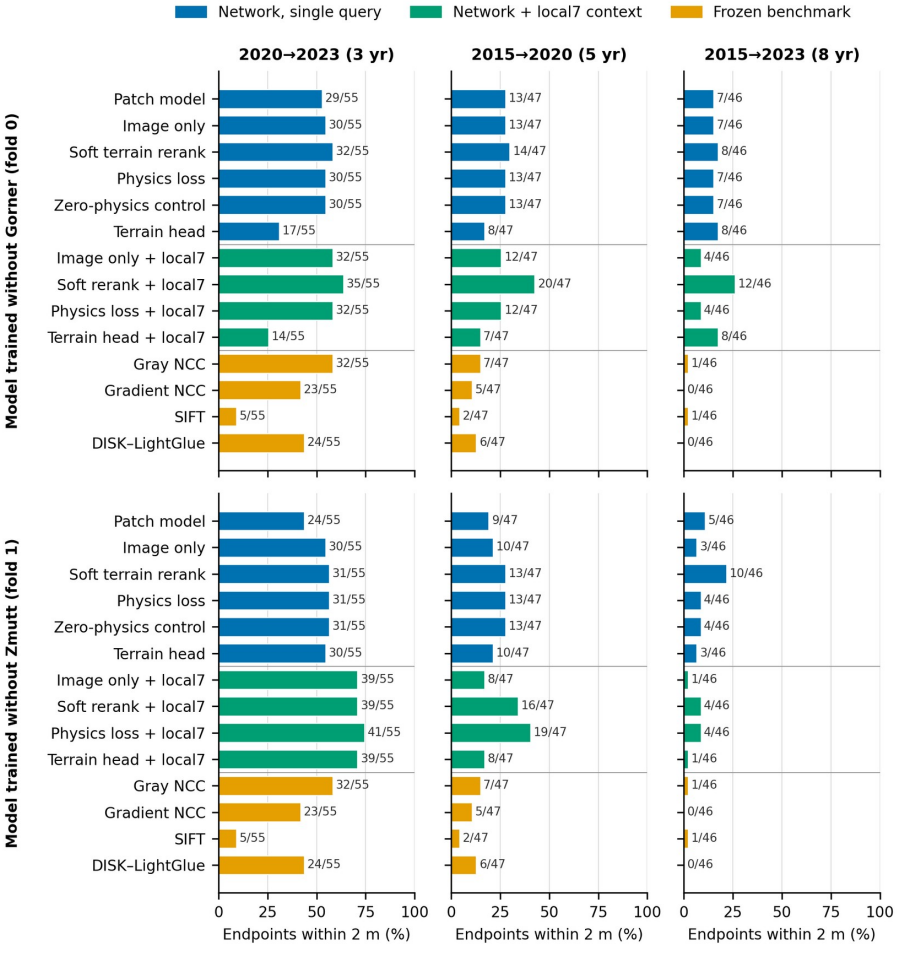


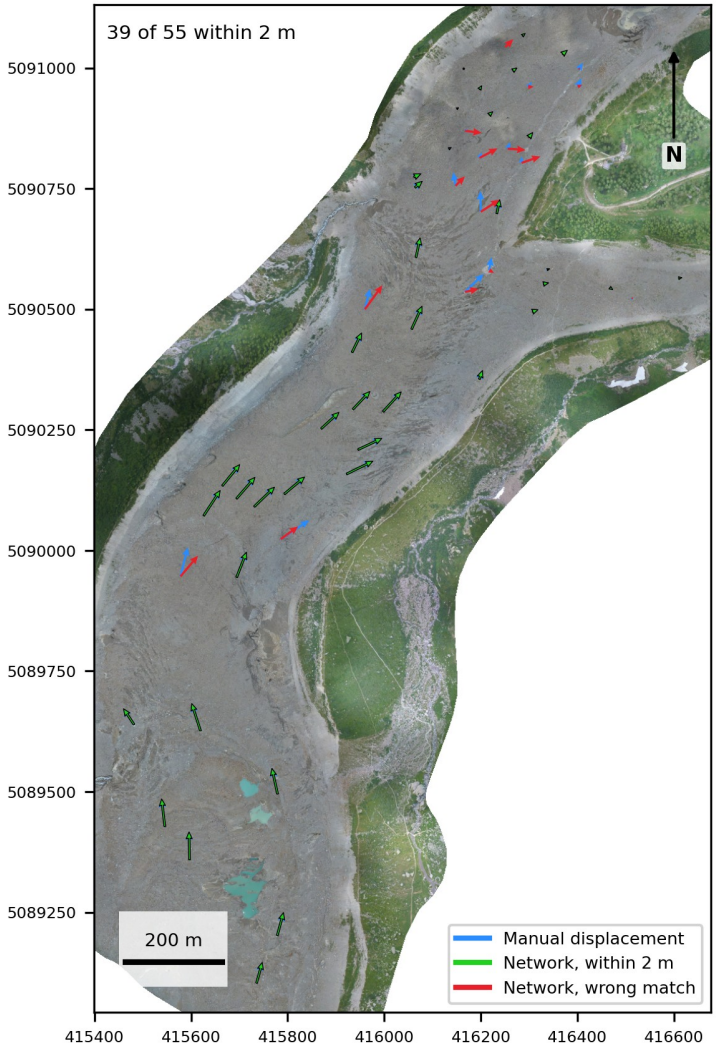
Figure 5. *Belvedere transfer test: share of boulder endpoints within 2 m of the manual reference for the network variants (blue: single query; green: with local7 context) and the frozen benchmarks (orange), for the three survey pairs (columns) and the two fold models (rows). Labels give successes/queries. Frozen benchmarks do not depend on the fold and are repeated in both rows.*

Soft reranking improved success over the image-only model in all six fold-and-interval combinations, by 1 to 7 queries, and never lost a boulder that the image-only model had found. Its largest gain was on 2015→2023 with the Zmutt-fold model, from 3/46 to 10/46 (Wilson interval 12–36 %), but the median error remained 101 m. In the Gorner fold the terrain head reduced success on 2020→2023 and 2015→2020 (17/55, 8/47) and gained one boulder on 2015→2023 (8/46); in the Zmutt fold it was identical to the image-only model.

Local context helped most on the 3-year pair, raising soft-rerank success to 35/55 (Gorner fold) and 39/55 (Zmutt fold). On longer pairs its effect was mixed. Combined with soft reranking it raised success on 2015→2020 in both folds (to 20/47 and 16/47) and on 2015→2023 in the Gorner fold (to 12/46), but lowered it on 2015→2023 in the Zmutt fold (from 10/46 to 4/46); image-only + local7 fell to 1/46 on that pair. The highest single figure on Belvedere, 41 of 55 (75 %), came from the physics-loss model with

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

277 local7 in the Zmutt fold. The physics-loss model and its zero-physics control returned the same endpoint
278 for 54 of 55 boulders on this pair, so the gain reflects the additional fine-tuning rather than the physical
279 term. We do not use it as the headline result, since that would mean picking the best of 16 variant-fold
280 combinations after seeing the scores.



281

282 ***Figure 6.*** *Belvedere Glacier, 2020→2023: manual displacement vectors (blue) and predicted vectors within 2 m of the manual*
283 *endpoint (green) or wrong matches (red) for all 55 query boulders, fold 1 model (trained without Zmutt) with soft terrain rerank*
284 *and local7 context; 39 of 55 endpoints within 2 m. Background: 2020 UAV orthophoto (0.2 m GSD; Ioli et al., 2024). Coordinates*
285 *are in EPSG 32632. Maps for all three pairs and both fold models are shown in Fig. S3.*

286 Among the frozen methods, gray NCC was the strongest on the 3-year pair (32/55), equal to the Gorner-
287 fold soft terrain rerank (32/55) and slightly above the image-only network (30/55 in both folds); the soft
288 rerank + local7 variants reached 35–39/55. Gradient NCC and DISK–LightGlue reached 23/55 and 24/55,

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

and SIFT returned an endpoint for only 10 of 55 queries. On the longer pairs all frozen methods dropped sharply, to at most 7/47 (2015→2020) and 1/46 (2015→2023), whereas the image-only and soft-rerank network variants kept 10–14/47 and 3–10/46, depending on fold.

3.3 Plausibility of predicted displacements

Soft reranking reduced the number of upstream or uphill predictions relative to the image-only model on all six Belvedere fold-and-interval sets; in the Zmutt fold it removed them almost entirely (at most one flag per pair). On held-out tests, upstream predictions for image-only, soft rerank, physics and terrain head were 3, 4, 3 and 0 of 13 on Gorner and 11, 0, 9 and 11 of 23 on Zmutt. The Gorner-fold terrain head produced no upstream or uphill flags on Belvedere, but its localization was worse or nearly unchanged. Fewer implausible directions therefore did not translate into more correct endpoints. Note that the manual reference itself contains 2 of 13 upstream cases on the Gorner test, so the audit is a consistency check, not ground truth.

3.4 Surface speed and flow regimes

Converted to surface speed, the manual tracks of 2020→2023 fall into two regimes (Fig. 7a): about half of the boulders move slower than 7 m yr^{-1} (median 3.4 m yr^{-1}) and one third faster than 17 m yr^{-1} (median 19.1 m yr^{-1}); for 2015→2020 the medians of the two classes are 5.4 and 19.6 m yr^{-1}. These values agree with the 2–7 m yr^{-1} and 17–22 m yr^{-1} reported for 2015–2020 by Ioli et al. (2022) for the slow and fast parts of the glacier, which supports the manual reference. Because Ioli et al. (2022) do not give sector coordinates, the comparison is between velocity regimes, not between identical areas. Where the network relocated a boulder within 2 m, its speed differed from the manual speed by a median of 0.11–0.17 m yr^{-1} (maximum 0.48 m yr^{-1}), comparable to the ±0.3 m yr^{-1} uncertainty of manual feature tracking on Belvedere (Ioli et al., 2022). Over all 55 boulders, wrong matches included, the median absolute speed difference was 0.13–0.17 m yr^{-1}.

With soft reranking and local context, the speed class of the network agreed with the manual class for 75 % (fold 0) and 84 % (fold 1) of the boulders over 2020→2023. Above all, not a single boulder was assigned to the opposite flow regime in either fold: no slow boulder was classed as fast and no fast boulder as slow, and all misclassified boulders fell into the intermediate class (Fig. 7c). This is the effect of local context. Without it, the same models classed 5–6 slow boulders per fold as fast, i.e. slow ice as fast-flowing ice. Most wrong matches lay in the slower northern part of the surveyed tongue (Fig. 7b). The wrong matches of the two folds also differ systematically (Fig. 7c): the model trained without Gorner tended to underestimate the speed of fast boulders, whereas the model trained without Zmutt tended to overestimate the speed of slow boulders. On the longer pairs, class agreement dropped to 57–66 % (2015→2020) and 46–57 % (2015→2023), but local context still limited regime inversions to at most one boulder in each direction per fold and pair, against up to 13 slow boulders classed as fast without it.

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

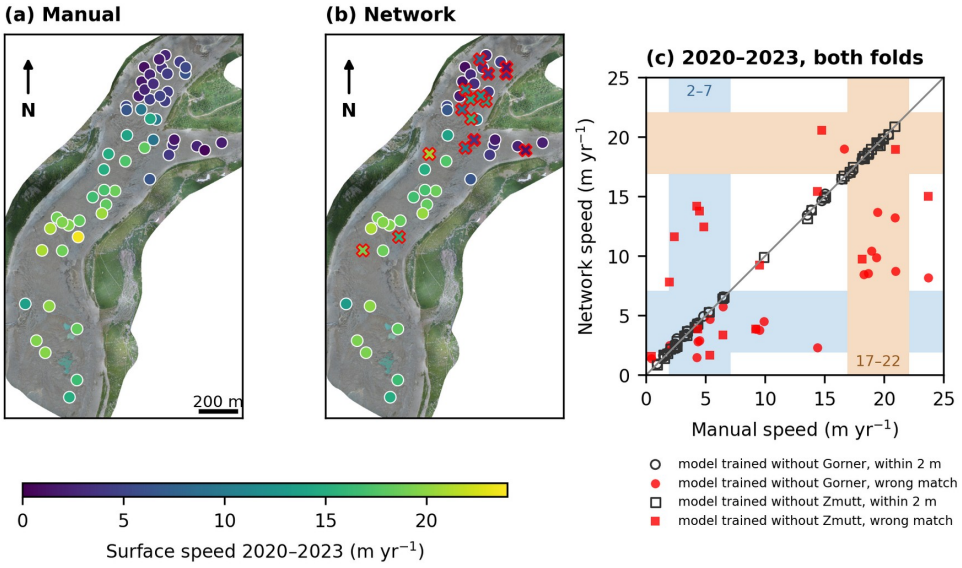


Figure 7. *Surface speed on Belvedere Glacier, 2020→2023, from boulder displacements; points are placed at the 2020 source position. (a) Manual tracks. (b) Fold 1 model (trained without Zmutt) with soft terrain rerank and local7 context; crosses mark endpoints more than 2 m from the manual position. (c) Network versus manual speed for both fold models; shaded bands show the 2–7 and 17–22 m yr^{-1} ranges reported for 2015–2020 by Ioli et al. (2022). Background: 2020 UAV orthophoto (Ioli et al., 2024).*

4 Discussion

4.1 Performance and practical use

On an unseen glacier and over a 3-year interval, the network with terrain reranking and local context relocates about two thirds of the query boulders within 2 m (64–71 %), and 56–58 % from single queries. The single-query result is on par with the best frozen benchmark tested (gray NCC, 58 %) and within the uncertainty of 55 queries; local context adds 5–15 percentage points and places the network 5–13 points above gray NCC. On longer intervals the margin is larger, because cross-correlation and the feature-based matchers tested lose almost all skill.

With the model trained without Zmutt, soft terrain reranking and local context, 39 of 55 boulders (71 %; Wilson 95 % interval 58–81 %) were relocated within 2 m over the 3-year interval 2020→2023, on a glacier never seen in training. The model trained without Gorner gives 35 of 55 (64 %; 50–75 %), so the result holds, with some variation, whichever glacier is left out of training. We take soft rerank + local7 as the main configuration because soft reranking is the only terrain variant that improved results consistently; it is not the best variant on every set (Sect. 3.2), and the use of neighboring predictions makes it an optimistic figure. The corresponding single-query results are 32 and 31 of 55 (58 % and 56 %). Even so, it is a notable result for a task that, to our knowledge, is still carried out largely by hand: from the photogrammetric surveys of Miage in the 1970s (Lesca, 1974a, b) to recent studies on rock glaciers and debris-covered glaciers (Scotti et al., 2017; Mölg et al., 2019; Ioli et al., 2022; Tanarro et al., 2026).

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

The motivating application is to tell flowing from stagnant ice on thinning debris-covered tongues. A 2 m endpoint tolerance over 3 years corresponds to about 0.7 m yr^{-1}, well below the 3 m yr^{-1} used to describe the quasi-stagnant lower tongue of Zmuttgletscher (Mölg et al., 2019). The problem is not precision but reliability. When the network fails, errors reach tens to hundreds of meters. An operational tool therefore needs a confidence score or an abstention rule, so that wrong matches are flagged rather than reported as velocity. The speed classes of Sect. 3.4 show that, with local context, the network never assigned a slow boulder to the fast regime, or a fast boulder to the slow regime, over three years. This is precisely the error that would turn slow, possibly stagnating ice into fast-flowing ice, or the reverse, in a map of flow regimes, and it is the information needed to detect a transition toward stagnation. The slow sector, however, is also where most wrong matches occurred, so slow ice does not make matching easier; whether this reflects a more chaotic debris surface or smaller and less distinctive boulders remains to be tested.

The opposite biases of the two folds (Sect. 3.4) mirror the dynamics of the glacier left out of training. The Gorner test boulders moved at about 14 m yr^{-1} over 1999–2005, the Zmutt test boulders at about 4 m yr^{-1} (Sect. 4.3). Without Gorner, the training set contains fewer large displacements, and wrong matches tend to fall short of fast boulders; without Zmutt, it contains fewer slow, near-stagnant boulders, and wrong matches tend to overshoot slow ones. A wrong match is the selection of another boulder rather than a biased speed estimate, so this pattern suggests that the network learns a prior on plausible displacements from its training glaciers. With 16–20 wrong matches per fold, this interpretation remains tentative, but it indicates that training glaciers should span the range of flow speeds expected at the target site.

4.2 Effect of survey interval and image quality

All methods lost most of their skill on the 8-year pair. Several effects probably combine. First, the boulder surface itself changes: boulders rotate, topple from ice cliffs and hummocks, and are partly buried or exposed as the debris layer redistributes. Second, the search area grows with the interval. At 17–22 m yr^{-1} (Ioli et al., 2022), 8 years of motion needs a radius near 200 m, which contains many similar-looking boulders. Third, the quality of the oldest orthophoto is lower. The 2015 Belvedere orthophoto was acquired on two dates two weeks apart, 8 and 23 October 2015, and the mosaic appears to combine four flights, two on each day, i.e. at different times of day. The resulting patches have markedly different illumination, so color, contrast and shadows change across the tongue and between neighboring boulders. The low autumn sun also casts long shadows, which alter the apparent shape of boulders. The later orthophotos were acquired in summer, on 26–27 July 2020 and 24–26 July 2023. The 2020 mosaic, although assembled from two products presumably acquired at different times, is homogeneous in illumination, but it shows smearing artifacts in several sectors, where small boulders retain little distinctive texture and may be matched to a different boulder (Fig. S4); the 2023 mosaic is very homogeneous. In the 2015 mosaic, parts of the tongue are also covered by fresh snow, which hides boulders entirely (Fig. 2): boulders under snow could not be digitized and were excluded, which explains most of the drop from 55 to 47 queries (Sect. 2.1), and boulders near the snow margin may be partly covered or shadowed. Both pairs that include 2015 (2015→2020 and 2015→2023) perform worse than 2020→2023 for every method, which is consistent with this effect, although it cannot be separated from the longer interval. Cross-correlation is known to fail under large displacements and poor image sharpness, which is why boulders on older regional orthophotos are still tracked by hand (Scotti et al., 2017). Human operators rely on the position of a boulder relative to its neighbors and to surface

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

landforms, the same kind of context that our local7 consensus exploits. Figure 8 separates the role of displacement from that of the survey pair. In practice, chaining shorter intervals is likely to be more reliable than direct long-interval matching. Our direct 2015→2023 pair deliberately did not chain through 2020.

For the 46 boulders tracked in all three Belvedere surveys, the median manual displacement grows from 18 m (2020→2023) to 36 m (2015→2020) and 48 m (2015→2023), while success of the soft-rerank model with local7 falls from 59–67 % to 33–41 % and 9–26 % (Fig. 8a). These values refer to the 46 common boulders; on all 55 boulders of 2020→2023 the same model reaches 64–71 % (Sect. 4.1). Within each pair, however, success does not decrease systematically with displacement (Fig. 8b). On 2020→2023, boulders that moved 50–100 m were relocated in 42–83 % of cases, against 65–73 % for those that moved less than 25 m and 38–50 % for 25–50 m. The longest displacements of the 8-year pair (100–170 m) were still found in 11–16 % of cases. Displacement length, and hence the size of the search area, is therefore not the main cause of the drop. The survey pair matters more: the pairs that include the October 2015 orthophoto perform worse, on average over the two folds, in almost every displacement class, consistent with the image-quality and snow effects described above and with the surface change that accumulates over longer intervals. The two fold models also differ markedly on the 8-year pair (12/46 and 4/46).

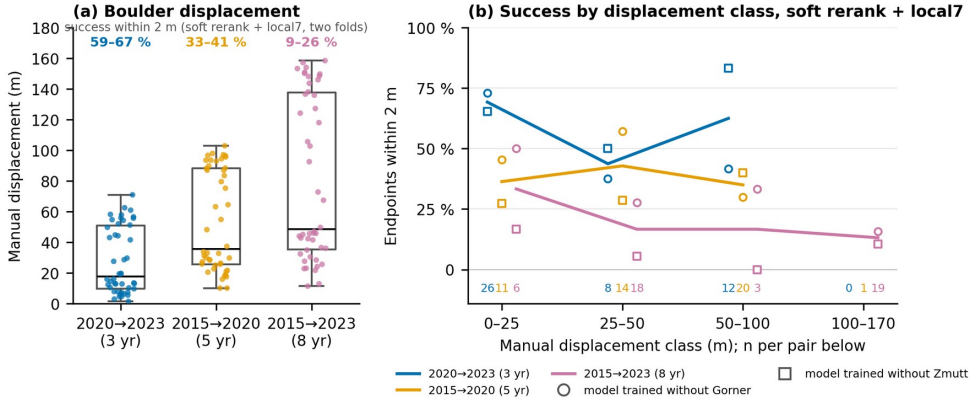


Figure 8. *(a) Manual displacement of the 46 boulders tracked in 2015, 2020 and 2023 on Belvedere, for each survey pair; numbers above give the share of endpoints within 2 m for the soft-rerank model with local7 context of the two folds. (b) Share of endpoints within 2 m by displacement class, same model; open symbols are the two fold models, lines their mean; numbers at the bottom give the boulders per class and pair.*

A coarse terrain prior helped modestly and consistently when used to rerank candidates, and it reduced upstream and uphill predictions. Adding it as a training loss had no detectable effect, and a learned terrain head was unstable. The manual reference itself contains upstream displacements (2 of 13 Gorner test boulders), which are physically possible where boulders slide off ice cliffs or hummocks; terrain information should therefore act as a soft preference rather than a hard constraint. Both global DEMs used here are far from contemporaneous with the imagery, Copernicus GLO-30 being derived from 2011–2015 acquisitions (ESA, 2021) and SRTMGL1 from February 2000 (NASA JPL, 2013), and both are too coarse to resolve the debris relief. They can nevertheless indicate the general down-glacier direction, which is the information the rerank uses.

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

4.3 Transferability and limitations

Belvedere was excluded from training, validation and hyperparameter selection, and its labels were read only after predictions were saved. The search radii were derived from training-glacier displacements and published Belvedere velocities, not from the Belvedere tracks (Sect. 2.5). In this sense the Belvedere results are a zero-shot transfer test. The two fold models are evaluated on the same Belvedere tracks, so their results are paired, not independent replicates.

The transfer is also a transfer across image types. The training data span a factor of ten in ground resolution and three acquisition platforms: UAV orthophotos at 0.05 m (Ciardoney), helicopter orthophotos at 0.15 m (Miage) and national aerial orthophotos at 0.50 m (Gorner, Zmutt). The model was then tested on aerial orthophotos at 0.50 m (held-out Gorner and Zmutt) and on UAV orthophotos at 0.20 m (Belvedere), a native resolution absent from the training imagery, although patch training simulates effective resolutions of 0.20 and 0.30 m by downsampling (Sect. 2.3). That a single model can be applied across this range, with a fixed 0.10 m processing grid, supports its use on the heterogeneous orthophoto archives available for most Alpine glaciers. Results on the two held-out glaciers differ markedly (at best 8 of 13 boulders within 2 m on Gorner, 18 of 23 on Zmutt), although both use the same 0.50 m swisstopo product. The Gorner test boulders moved much farther over 1999–2005 (median 84 m, range 57–153 m, i.e. about 14 m yr^{-1}) than the Zmutt test boulders (median 25 m, range 4–67 m, about 4 m yr^{-1}). Faster flow implies more surface change between surveys and endpoints closer to the edge of the 160 m search window.

The composition of the training set should be kept in mind when interpreting the Belvedere results. Miage contributes 184 of the 300 (fold 0) and 281 (fold 1) training queries, about 61–65 %, and its large debris-covered tongue is morphologically similar to that of Belvedere. Part of the transfer skill may therefore reflect this similarity. The model is, however, not trained on debris-covered glaciers alone: Ciardoney, a small glacier that is still largely debris-free and only slowly acquiring a debris cover, contributes 71 queries (about 24–25 %), and Gorner and Zmutt add debris-covered tongues imaged at a much coarser resolution. Testing on a debris-covered glacier with a morphology different from Miage, and on further clean-ice glaciers, would show how much the result depends on this similarity.

Test sets are small (13 and 23 held-out queries), and the design has only two folds. The validation and test halves of each held-out glacier are split along its principal spatial axis; their labels are separate, but their image and search areas may overlap. As in other studies based on manual tracks, the reference displacements also include the co-registration and digitization errors of the orthophotos, which were not quantified separately here.

5 Conclusions

We trained a correspondence network to reproduce manual supraglacial boulder tracking from orthophotos, with optional coarse terrain information and a consensus among neighboring boulders, and tested it on held-out glaciers and on an unseen glacier. Over a 3-year interval on Belvedere, with terrain reranking and local context, it relocates 64–71 % of the boulders within 2 m (56–58 % from single queries, on par with the best of the four frozen benchmarks tested); over 5 and 8 years the network keeps a clear margin over the frozen benchmarks, which lose almost all skill. Converted to surface speed, correct matches reproduce the manual speeds within 0.5 m yr^{-1}, and, with local context, no boulder is assigned to the opposite flow regime over three years. For a task that is still carried out

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

largely by hand, this makes semi-automatic tracking realistic: the operator verifies the proposals and re-tracks about one boulder in three. Over 8 years all methods fail for most boulders, also because of lower image quality and snow in the oldest survey. A soft terrain prior and local context improved the results, whereas the physics-motivated training loss, in the form tested here, brought no measurable benefit. Automated boulder tracking is therefore realistic for short, regularly spaced surveys, which is how UAV and regional orthophoto programs are increasingly acquired.

Code availability

The code used in this study is available at https://github.com/Azoso95/glacier-boulder-tracking (commit d060d1b) under the GNU General Public License v3.0. A version archived on Zenodo with a DOI will be provided with the revised manuscript.

Data availability

The Belvedere UAV orthophotos are openly available on Zenodo (Ioli et al., 2024), and the swisstopo aerial images used for Gorner and Zmutt are available from swisstopo (swisstopo, 2026a). Glacier outlines are from RGI 7.0 (RGI Consortium, 2023), and the DEMs from the Copernicus GLO-30 (ESA, 2021) and SRTMGL1 (NASA JPL, 2013) products. The Miage and Ciardoney orthophotos, the manual boulder tracks of all glaciers and the model outputs are available from the corresponding authors upon request.

Author contributions

FP and AF contributed equally to this work. FP: conceptualization, data curation (digitization and checking of all boulder reference tracks), investigation, formal analysis, visualization, writing - original draft. AF: methodology, software (tracking network and processing pipeline), validation. LP, MG: supervision, review. All authors discussed the results and approved the final manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We thank Walter Alberto (Fondazione Glaciologica Italiana) for the first digitization of the Miage boulder tracks. We thank the Società Meteorologica Italiana (SMI) for making available the Ciardoney orthophotos acquired over years of collaboration.

LLMs were used solely for textual corrections and code debugging. The authors retain full responsibility for the ideas, results, and interpretations presented in the manuscript.

References

- Benoit, L., Gourdon, A., Vallat, R., Irarrazaval, I., Gravey, M., Lehmann, B., Prasicek, G., Gräff, D., Herman, F., and Mariethoz, G.: A high-resolution image time series of the Gorner Glacier - Swiss Alps - derived from repeated unmanned aerial vehicle surveys, Earth Syst. Sci. Data, 11, 579–588, https://doi.org/10.5194/essd-11-579-2019, 2019.

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

- Carniel, N., Tabernig, R., Hartl, L., and Höfle, B.: Deriving activity zones of the Äußeres Hochebenkar rock glacier through boulder tracking in multitemporal UAV-LiDAR point clouds, EGU General Assembly 2025, Vienna, Austria, 27 April–2 May 2025, EGU25-5020, https://doi.org/10.5194/egusphere-egu25-5020, 2025.
- Cogley, J. G., Hock, R., Rasmussen, L. A., Arendt, A. A., Bauder, A., Braithwaite, R. J., Jansson, P., Kaser, G., Möller, M., Nicholson, L., and Zemp, M.: Glossary of Glacier Mass Balance and Related Terms, IHP-VII Technical Documents in Hydrology No. 86, IACS Contribution No. 2, UNESCO-IHP, Paris, 2011.
- ESA: Copernicus DEM - Global and European Digital Elevation Model (COP-DEM), GLO-30, https://doi.org/10.5270/ESA-c5d3d65, 2021.
- Hartl, L., Zieher, T., Bremer, M., Stocker-Waldhuber, M., Zahs, V., Höfle, B., Klug, C., and Cicoira, A.: Multi-sensor monitoring and data integration reveal cyclical destabilization of the Äußeres Hochebenkar rock glacier, Earth Surf. Dynam., 11, 117–147, https://doi.org/10.5194/esurf-11-117-2023, 2023.
- Hartmeyer, I., Delleske, R., Keuschnig, M., Krautblatter, M., Lang, A., Schrott, L., and Otto, J.-C.: Current glacier recession causes significant rockfall increase: the immediate paraglacial response of deglaciating cirque walls, Earth Surf. Dynam., 8, 729–751, https://doi.org/10.5194/esurf-8-729-2020, 2020.
- Heid, T. and Kääb, A.: Evaluation of existing image matching methods for deriving glacier surface displacements globally from optical satellite imagery, Remote Sens. Environ., 118, 339–355, https://doi.org/10.1016/j.rse.2011.11.024, 2012.
- Hendrickx, H., Elias, M., Blanch, X., Delaloye, R., and Eltner, A.: AI-based tracking of fast-moving alpine landforms using high-frequency monoscopic time-lapse imagery, Earth Surf. Dynam., 13, 705–721, https://doi.org/10.5194/esurf-13-705-2025, 2025.
- Herreid, S. and Pellicciotti, F.: The state of rock debris covering Earth's glaciers, Nat. Geosci., 13, 621–627, https://doi.org/10.1038/s41561-020-0615-0, 2020.
- Ioli, F., Bianchi, A., Cina, A., De Michele, C., Maschio, P., Passoni, D., and Pinto, L.: Mid-term monitoring of glacier's variations with UAVs: the example of the Belvedere Glacier, Remote Sens., 14, 28, https://doi.org/10.3390/rs14010028, 2022.
- Ioli, F., De Gaetani, C., Barbieri, F., Gaspari, F., Pinto, L., and Rossi, L.: Belvedere Glacier long-term monitoring Open Data - Pointclouds, DEMs, Orthophotos, Version 2.0, Zenodo, https://doi.org/10.5281/zenodo.10817029, 2024.
- Kaushik, S., Singh, T., Bhardwaj, A., Joshi, P. K., and Dietz, A. J.: Automated delineation of supraglacial debris cover using deep learning and multisource remote sensing data, Remote Sens., 14, 1352, https://doi.org/10.3390/rs14061352, 2022.
- Kaushik, S., Maurya, L., Tellman, E., Zhang, G., and Dharpure, J. K.: Debris covered glacier mapping using newly annotated multisource remote sensing data and geo-foundational model, Sci. Remote Sens., 12, 100319, https://doi.org/10.1016/j.srs.2025.100319, 2025.
- Lesca, C.: Emploi de la photogrammétrie analytique pour la détermination de la vitesse superficielle des glaciers et des profondeurs relatives, Boll. Comit. Glac. It., 22, 169–186, 1974a.
- Lesca, C.: Méthode topographique de précision pour la détermination de la vitesse superficielle des glaciers, Boll. Comit. Glac. It., 22, 153–168, 1974b.

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.

- Lindenberger, P., Sarlin, P.-E., and Pollefeys, M.: LightGlue: local feature matching at light speed, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 17627–17638, https://doi.org/10.1109/ICCV51070.2023.01616, 2023.
- Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S.: A ConvNet for the 2020s, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 11976–11986, https://doi.org/10.1109/CVPR52688.2022.01167, 2022.
- Lowe, D. G.: Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision, 60, 91–110, https://doi.org/10.1023/B:VISI.0000029664.99615.94, 2004.
- Mölg, N., Bolch, T., Walter, A., and Vieli, A.: Unravelling the evolution of Zmuttgletscher and its debris cover since the end of the Little Ice Age, The Cryosphere, 13, 1889–1909, https://doi.org/10.5194/tc-13-1889-2019, 2019.
- NASA JPL: NASA Shuttle Radar Topography Mission Global 1 arc second, Version 003, NASA EOSDIS Land Processes Distributed Active Archive Center, https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003, 2013.
- Natural Earth: Admin 0 - Countries, 1:110m, version 5.1.1, https://www.naturalearthdata.com, last access: 8 October 2026.
- Norova, N., Samat, A., and Abuduwaili, J.: A systematic review of methodological advances in glacier-velocity retrieval with an emphasis on debris-covered glaciers, Remote Sens., 18, 62, https://doi.org/10.3390/rs18010062, 2026.
- Pope, A.: Glacier or not? The importance of nuance in definitions of vanishing glaciers, Ann. Glaciol., 66, e32, https://doi.org/10.1017/aog.2025.10030, 2025.
- Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., and Feichtenhofer, C.: SAM 2: Segment anything in images and videos, [preprint], https://doi.org/10.48550/arXiv.2408.00714, 2024.
- RGI Consortium: Randolph Glacier Inventory - A Dataset of Global Glacier Outlines, Version 7 (NSIDC-0770), NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA, https://doi.org/10.5067/f6jmovy5navz, 2023.
- Scotti, R., Crosta, G. B., and Villa, A.: Destabilisation of creeping permafrost: the Plator rock glacier case study (Central Italian Alps), Permafrost Periglac. Process., 28, 224–236, https://doi.org/10.1002/ppp.1917, 2017.
- Shankar, S., Stearns, L. A., and van der Veen, C. J.: Semantic segmentation of glaciological features across multiple remote sensing platforms with the Segment Anything Model (SAM), J. Glaciol., 70, e4, https://doi.org/10.1017/jog.2023.95, 2024.
- Stefaniak, A., Robson, B. A., Cook, S. J., Clutterbuck, B., Midgley, N. G., and Labadz, J. C.: Mass balance and surface evolution of the debris-covered Miage Glacier, 1990–2018, Geomorphology, 373, 107474, https://doi.org/10.1016/j.geomorph.2020.107474, 2021.
- swisstopo: Aerial images swisstopo colour (LUBIS), dataset ch.swisstopo.lubis-luftbilder_farbe, Federal Office of Topography swisstopo, Wabern, Switzerland, https://opendata.swiss/en/dataset/luftbilder-swisstopo-farbig, last access: 5 October 2026, 2026a.
- swisstopo: SWISSIMAGE 10 cm, Federal Office of Topography swisstopo, Wabern, Switzerland, https://www.swisstopo.admin.ch/en/orthoimage-swissimage-10, last access: 27 September 2026, 2026b.

https://doi.org/10.5194/egusphere-2026-6731
Preprint. Discussion started: 24 November 2026
© Author(s) 2026. CC BY 4.0 License.



- Tanarro, L. M., Palacios, D., Fernández-Fernández, J. M., Santos-González, J., González-Gutiérrez, R. B., Andrés, N., Schimmelpfennig, I., Farnsworth, W. R., Melón-Nava, A., Peña-Pérez, S. A., Brynjólfsson, S., Sæmundsson, Þ., and ASTER Team: Unveiling the evolution of a glacier through geomorphological analysis and surface exposure dating: Nautárdalsjökull (Northern Iceland), CATENA, 268, 109981, https://doi.org/10.1016/j.catena.2026.109981, 2026.
- Tyszkiewicz, M. J., Fua, P., and Trulls, E.: DISK: learning local features with policy gradient, in: Advances in Neural Information Processing Systems 33 (NeurIPS 2020), Virtual, 6–12 December 2020, 14254–14265, 2020.
- Vivero, S., Hendrickx, H., Frankl, A., Delaloye, R., and Lambiel, C.: Kinematics and geomorphological changes of a destabilising rock glacier captured from close-range sensing techniques (Tsarmine rock glacier, Western Swiss Alps), Front. Earth Sci., 10, 1017949, https://doi.org/10.3389/feart.2022.1017949, 2022.
- Zandler, H., Abermann, J., Robson, B. A., Maschler, A., Scheiber, T., Carrivick, J. L., and Yde, J. C.: Deep learning outperforms existing algorithms in glacier surface velocity estimation with high-resolution data - the example of Austerdalsbreen, Norway, Front. Remote Sens., 6, 1586933, https://doi.org/10.3389/frsen.2025.1586933, 2025.