Towards Reliable Sea Ice Drift Estimation in the Arctic: Deep Learning Optical Flow on RADARSAT-2

Daniela Martin
University of Delaware
dmartinv@udel.edu

Joseph Gallego
Drexel University
jg3959@drexel.edu

Abstract

Accurate estimation of sea ice drift is critical for Arctic navigation, climate research, and operational forecasting. While *optical flow*a computer vision technique for estimating pixel-wise motion between consecutive imageshas advanced rapidly in computer vision, its applicability to geophysical problems, and more specifically, to satellite SAR imagery, remains underexplored. Classical optical flow methods rely on mathematical models and strong assumptions about motion, which limit their accuracy in complex scenarios; recent deep learning-based approaches have substantially improved performance and are now the standard in computer vision, motivating their application to sea ice drift estimation. We present the first large-scale benchmark of 48 deep learning optical flow models on RADARSAT-2 ScanSAR sea ice imagery, evaluated with endpoint error (EPE) and Fl-all metrics against GNSS-tracked buoys. Several models achieve sub-kilometer accuracy (EPE 6-8 pixels, 300-400 m), a small error relative to the spatial scales of sea ice motion and typical navigation requirements in the Arctic. Our results demonstrate that the models are capable of capturing consistent regional drift patterns and that recent deep learning-based optical flow methods, which have substantially improved motion estimation accuracy compared to classical methods, can be effectively transferred to polar remote sensing. Optical flow produces spatially continuous drift fields, providing motion estimates for every image pixel rather than at sparse buoy locations, offering new opportunities for navigation and climate modeling.

1 Introduction

Arctic sea ice is a key component of the climate system, influencing global circulation patterns [19, 40], ecosystems [32, 5], and human activity through its role in navigation and forecasting [44, 3]. Accurate estimation of ice drift is therefore essential for both climate research and operational decision-making [36]. Classical approaches for sea ice motion estimation include correlation-based methods such as maximum cross-correlation and phase correlation [17, 12], as well as feature tracking strategies [16]. While effective in certain conditions, these methods often degrade in low-texture regions, at floe boundaries, or under strong deformation [28, 18].

Optical flow is a computer vision technique that estimates motion at the level of individual pixels between two consecutive images. Each pixel is assigned a vector indicating the direction and magnitude of its movement from the first to the second image, providing a dense motion field across the entire scene [1]. In the context of sea ice, optical flow has the potential to capture detailed drift patterns over large areas, including subtle deformations and floe interactions, which are difficult to resolve with sparse tracking methods such as drifting buoys, whose limited spatial coverage prevents the observation of fine-scale drift patterns and interactions between individual ice floes. Classical optical flow approaches rely on mathematical models, such as brightness constancy constraints combined

with spatial regularization, solved through variational or local optimization methods [39]. While effective in certain contexts, these methods are limited in handling large displacements, complex motion patterns, or noisy data. In recent years, deep learning has become the standard for computer vision tasks, including optical flow, enabling models to learn motion patterns from data and achieve state-of-the-art accuracy and efficiency [31].

Recent developments in optical flow estimation have focused on improving both accuracy and efficiency, extending the RAFT [34] family of methods into more scalable and generalizable architectures. WAFT [37] challenges the need for cost volumes by relying on high-resolution warping, achieving competitive accuracy with lower memory usage and faster inference. MEMFOF [2] revisits RAFT-like designs with reduced correlation volumes and multi-frame estimation, striking a balance between runtime efficiency and state-of-the-art accuracy, particularly at FullHD resolution. DPFlow [26] addresses scalability by generalizing to ultra-high-resolution inputs (up to 8K) without requiring high-resolution training, while also introducing a new benchmark to systematically evaluate methods at large scales. Complementarily, SEA-RAFT [38] improves RAFTs convergence and generalization through a simplified architecture and novel training strategies, achieving strong cross-dataset performance with reduced computational cost. These approaches consistently achieve state-of-the-art results across standard optical flow benchmarks [21, 4, 9, 23]. Collectively, they highlight a shift toward architectures that not only improve accuracy but also address the growing demands of resolution, efficiency, and generalization in optical flow.

Progress in this direction has been limited in part by the scarcity of large-scale labeled datasets for Arctic sea ice motion. The 2021 *Sea Ice Dynamics Experiment (SIDEx)* campaign [14, 7] provides a rare opportunity by deploying a set of Global Navigation Satellite System (GNNS) [10] tracked buoys in the Arctic pack ice, delivering high-quality ground-truth trajectories that enable rigorous benchmarking of sea ice motion estimation at scale.

In this work, we present the first large-scale benchmark of 48 pretrained deep learning optical flow models, covering convolutional, recurrent, and transformer-based architectures, applied to RADARSAT-2 ScanSAR imagery [30] over the Beaufort Sea in the Alaskan Arctic. Models performance is evaluated against the buoy positions using widely adopted metrics in the computer vision community for optical flow assessment, specifically endpoint error (EPE) and Fl-all metrics. Several achieve sub-kilometer accuracy, highlighting the promise of pretrained vision models for polar geophysical applications. Beyond quantitative scores, we report qualitative patterns and discuss implications for Arctic navigation and climate science.

2 Experimental Setup

We evaluate the capability of deep learning optical flow methods to estimate sea ice drift from SAR imagery. Our experiments combine a dedicated RADARSAT-2 dataset with GNSS buoy observations, testing a wide range of pretrained models under consistent preprocessing and standardized evaluation metrics. This section outlines the dataset preparation, model selection, evaluation criteria, and quantitative and qualitative results.

2.1 Dataset Preprocessing

We use a dataset of 54 RADARSAT-2 ScanSAR images collected during the 2021 SIDEx campaign in the Beaufort Sea (March–May). The data provide wide coverage of Arctic sea ice at 20–50 m spatial resolution, encompassing diverse conditions such as compact ice, leads, and fractures. The temporal resolution of the dataset is $\approx$1.3 days.

For validation, we compare the estimated motion fields against ground-truth displacements from 43 buoys [14], including 12 SATICE buoys equipped with geodetic-quality GNSS tracking [7]. The SATICE buoys recorded positions every 5 minutes, while the other buoys provided updates every 10 minutes, enabling high-temporal-resolution tracking. To align with the SAR acquisitions, we linearly interpolated buoy positions, yielding 1880 displacement observations for evaluation. To ensure consistent model input, all images were resampled to a uniform 50 m spatial resolution. Each image was then cropped to a 40 km^2 region centered on the position of the GNSS buoy *OSU-IT-23* [14]. This Lagrangian frame of reference [8] allows the image patches to move with the buoy, retaining fine-scale deformation patterns.

2.2 Optical Flow Models

We benchmarked 48 pretrained optical flow models, spanning convolutional, recurrent, and transformer-based architectures. The selection was guided by both code availability and reproducibility of state-of-the-art optical flow methods reported in standard benchmarks [21, 4, 9, 23]. While some top-performing models have been reported in the literature, not all had accessible or verifiable implementations. To enable consistent evaluation with our sea ice data, we included only those models for which a reliable implementation could be obtained, using pretrained weights and interfaces provided by the *ptlflow* library [24].

Models were applied without fine-tuning to assess their transferability to SAR imagery. While our GNSS-tracked buoy observations provide valuable ground-truth for validation, they are too sparse-i.e., they provide measurements only at a limited number of discrete locations rather than at every pixel-to reliably train or fine-tune deep learning optical flow models. Most state-of-the-art optical flow architectures are trained on dense synthetic ground-truth displacement fields, which provide per-pixel motion vectors covering nearly the entire image [1], whereas our sea ice dataset contains only a handful of buoy-based measurements. Attempting to fine-tune on this sparse data would risk overfitting to the few available observations and would also reduce the amount of data available for an independent evaluation, since a subset would have to be withheld for validation [27]. By applying pretrained models directly, we can rigorously evaluate their out-of-the-box performance in a realistic Arctic scenario while preserving the integrity of the test set and avoiding spurious conclusions due to extremely limited training data.

Performance was quantified using two standard optical flow metrics: **Endpoint Error (EPE)**, the Euclidean distance between predicted and reference displacements (reported in pixels and converted to meters), and **Flow outlier rate (Fl-all)**, the proportion of pixels with error > 3 pixels and $> 5\%$ relative displacement. These metrics capture both absolute error and robustness to large displacements, providing complementary perspectives on model performance.

3 Results and Discussions

Quantitative Analysis. 75% of the tested models achieved EPE values within 6–8 pixels, indicating consistent performance across the majority of models. Table 1 presents the top five models, ranked by Fl-all to emphasize robustness against gross displacement errors in operationally relevant regions. Full quantitative results for all 48 models are provided in the supplemental material.

Table 1: Top-5 optical flow models evaluated on RADARSAT-2 ScanSAR imagery. Models are ranked by Fl-all, which quantifies the fraction of pixels with gross displacement errors, prioritizing reliability in operationally sensitive regions. EPE is reported in pixels (50 m/pixel). Lower values indicate better performance in both average error and robustness against gross errors. Best value for each metric is highlighted in bold.

Model	EPE [px]	Fl-all [%]
DIP (Sintel) [48]	6.29	**42.52**
RPKNet (Sintel) [25]	7.51	42.89
SEA-RAFT (M) (Spring) [38]	7.56	43.07
DPFlow (Chairs) [26]	**6.27**	43.08
SEA-RAFT (S) (Spring) [38]	7.75	43.12

Figure 1 presents a scatterplot of the best performing 22 optical flow models, showing EPE versus Fl-all. To improve readability, we only visualize models within a restricted range of errors. Specifically, we plot models with end-point error (EPE) below 8.5 and flow outlier rate (Fl-all) below 47. This filtering allows us to highlight the most competitive methods, since displaying all 48 evaluated models would overcrowd the figure and obscure relevant patterns.

Qualitative Analysis. To further assess model performance, we conducted a qualitative analysis on a representative RADARSAT-2 ScanSAR image pair. Figure 2 illustrates the predictions from the three best-performing models. It can be interpreted by referring to the standard optical flow colorwheel (subfigure c), where hue indicates direction and saturation indicates magnitude.
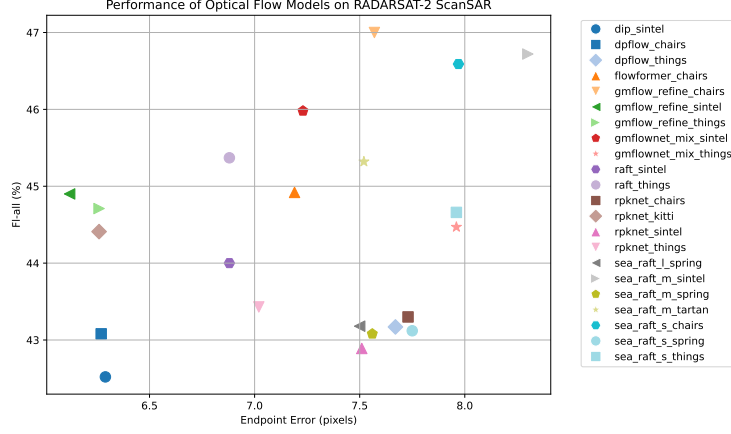
Figure 1: Scatter plot of optical flow models evaluated on RADARSAT-2 ScanSAR imagery, comparing end-point error (EPE) and flow outlier rate (Fl-all). For clarity, only models with EPE < 8.5 and Fl-all < 47 are shown (22 models in total). A total of 48 models were evaluated (full list in Table 2), but the full set is omitted here to avoid overcrowding. For comparison, the benchmark metrics officially reported in the PTLFlow library for standard computer vision datasets are summarized in Table 3.

Across all models (df), several distinct drift blocks can be identified, each characterized by a dominant color representing a coherent motion direction. The magnitude of motion increases gradually toward the corners of the scene and is lowest near the image center, an expected pattern given that the analysis is performed in a Lagrangian reference frame. All three models show a strong agreement in motion direction, capturing consistent regional drift pathways; however, differences emerge in the estimated magnitude.

The RPKNet model (e) systematically underestimates the motion magnitude compared to DIP (d) and SEA-RAFT (f), particularly in areas where multiple motion blocks converge. In addition, RPKNet produces spurious high-gradient values along the image borders that do not correspond to physically feasible ice motion. For practical use, we recommend cropping 5–10 pixels along the borders when applying this model to operational scenarios. In contrast, DIP and SEA-RAFT yield more stable and spatially coherent results, better preserving the large-scale drift structure.

For Arctic applications such as navigation support and early-warning systems, models that produce less nuanced and more categorical motion estimates—favoring stable regional drift patterns over subtle local variations—may be more reliable and interpretable. Overall, the qualitative results confirm that deep learning optical flow models can capture realistic sea ice motion fields, while highlighting the need for model-specific considerations depending on the intended operational context.

Discussion. Our large-scale benchmark of 48 deep learning optical flow models on RADARSAT-2 ScanSAR data demonstrates that pretrained computer vision methods can be effectively transferred to polar remote sensing. Remarkably, 75% of models achieved endpoint errors (EPE) within 6-8 pixels (approximately 300-400 m), with several methods surpassing this level of accuracy. Notably, Fl-all scores are relatively high because this metric is extremely strict, designed for contexts requiring subpixel precision; nevertheless, we focused on selecting the best values to minimize large errors as much as possible. While such errors may seem substantial in a general computer vision context, at Arctic scales they are negligible, since sea ice drift patterns evolve over tens of kilometers, making pixel-level deviations minor in the context of regional dynamics. The consistency across diverse architectures suggests that adaptations designed for complex textures and varying illumination in natural images also translate well to SAR imagery.

When compared to standard computer vision benchmarks (Table 3), the EPE values obtained for RADARSAT-2 imagery are higher in absolute terms. However, this discrepancy is expected given the fundamental differences between the domains: SAR imagery exhibits distinct speckle noise patterns and limited texture gradients, ground-truth motion fields are sparse and less precise, and the physical scales of motion are orders of magnitude larger. Therefore, while the reported errors
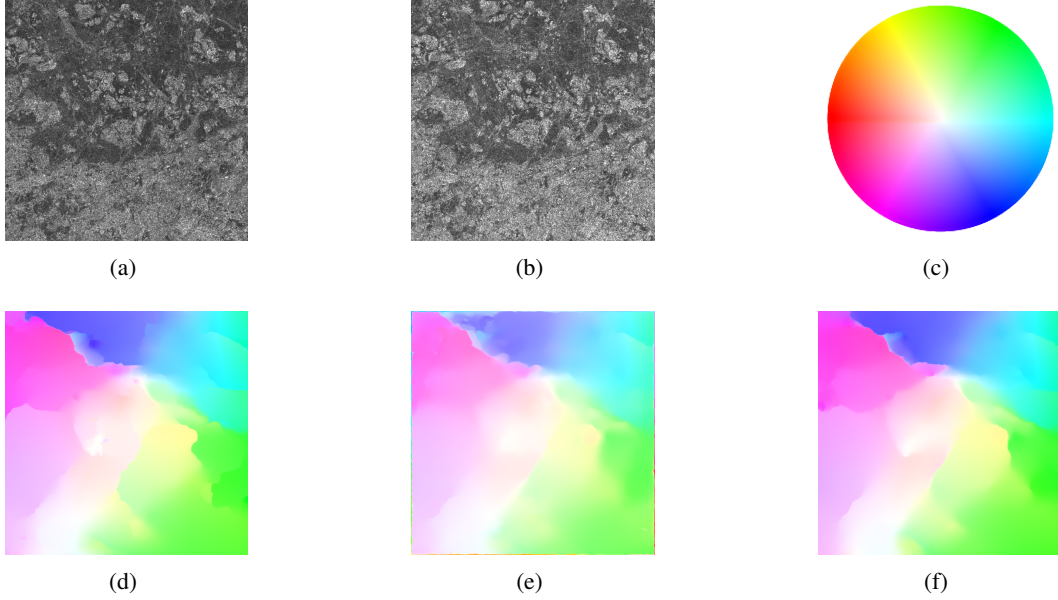
Figure 2: Qualitative comparison of the top three models on a selected RADARSAT-2 ScanSAR image pair. (a) First image. (b) Second image. (c) Colorwheel for interpretation: hue indicates direction, saturation indicates magnitude. Magnitude increases radially from the center (0) to the maximum displacement in the dataset ($\approx$100 pixels for this case) (d) DIP (Sintel) prediction [48]. (e) RPKNet (Sintel) prediction [25]. (f) SEA-RAFT (M) (Spring) prediction [38].

may seem high relative to benchmarks such as Sintel or KITTI, they are operationally negligible in Arctic applications, where ice drift occurs over tens of kilometers.

Qualitative analysis further showed that all models largely agree on motion patterns, although they may differ in edge definition and slightly in motion magnitude, while the motion direction is highly consistent. These qualitative observations further highlight subtle model-dependent trade-offs between smoothness, sharpness, and magnitude estimation that are consistent with the quantitative metrics reported above.

Limitations. A key limitation remains the lack of large, labeled SAR datasets for supervised training and fine-tuning. Manually generating ground truth fields at pixel level is nearly infeasible, making pretrained models a key ally in the near term.

4 Conclusion

Our benchmark provides the first comprehensive evidence that pretrained deep learning optical flow models can deliver operationally meaningful accuracies for sea ice drift estimation from SAR. These findings demonstrate the promise of transferring vision-based methods to geophysical applications and chart a path forward for developing domain-specific models tailored to the unique challenges of polar remote sensing.

Minor qualitative differences between architectures suggest that targeted adaptations to SAR-specific characteristics could further enhance accuracy and stability. Future progress will likely require domain-adapted approaches that combine machine learning with physical constraints, exploit multi-sensor fusion, and extend to long temporal sequences for seasonal-scale monitoring. Moreover, future work could explore the use of alternative evaluation metrics better suited to geophysical problems that capture spatial and temporal dependencies more effectively.

Acknowledgements

References

[1] Alfarano , A., Maiano , L., Papa , L., & Amerini , I. (2024) Estimating optical flow: A comprehensive review of the state of the art. *Computer Vision and Image Understanding* page 104160.

[2] Bargatin , V., Chistov , E., Yakovenko , A., & Vatolin , D. (2025) Memfof: High-resolution training for memory-efficient multi-frame optical flow estimation. *arXiv preprint arXiv:2506.23151*

[3] Bayler , E., Chang , P. S., De La Cour , J. L., Helfrich , S. R., Ignatov , A., Key , J., Lance , V., Leuliette , E. W., Byrne , D. A., Liu , Y., & others (2024) Satellite oceanography in noaa: Research, development, applications, and services enabling societal benefits from operational and experimental missions. *Remote Sensing* **16**(14).

[4] Butler , D. J., Wulff , J., Stanley , G. B., & Black , M. J. (2012) A naturalistic open source movie for optical flow evaluation. In A. Fitzgibbon et al. (Eds.), editor, *European Conf. on Computer Vision (ECCV)* pages 611–625. Springer-Verlag.

[5] Campbell , K., Matero , I., Bellas , C., Turpin-Jelfs , T., Anhaus , P., Graeve , M., Fripiat , F., Tranter , M., Landy , J. C., Sanchez-Baracaldo , P., & others (2022) Monitoring a changing arctic: Recent advancements in the study of sea ice microbial communities. *Ambio* **51**(2):318–332.

[6] Dosovitskiy , A., Fischer , P., Ilg , E., Hausser , P., Hazirbas , C., Golkov , V., Van Der Smagt , P., Cremers , D., & Brox , T. (2015) Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision* pages 2758–2766.

[7] Elosegui , P., Mondal , D., Ruszczyk , C., EcKert , C., Wilson , K., Hutchings , J., Bliss , A., Polashenski , C., Mahoney , A., Fedders , E., May , M., Manganini , K., Fitzgerald , C., & Forster , K. (2023) Sea ice drift tracks for the sea ice dynamic experiment (sidex) field campaign from geodetic global navigation satellite systems (gnss) satice buoys, alaska, 2021. *Arctic Data Center*

[8] Flotho , P., Heiss , C., Steidl , G., & Strauss , D. J. (2023) Lagrangian motion magnification with double sparse optical flow decomposition. *Frontiers in Applied Mathematics and Statistics* **9**:1164491.

[9] Geiger , A., Lenz , P., & Urtasun , R. (2012) Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*

[10] Grewal , M. S. (2011) Global navigation satellite systems. *Wiley Interdisciplinary Reviews: Computational Statistics* **3**(4):383–384.

[11] Haarpaintner , J., Tonboe , R. T., Long , D. G., & Van Woert , M. L. (2004) Automatic detection and validity of the sea-ice edge: An application of enhanced-resolution quikscat/seawinds data. *IEEE Transactions on Geoscience and Remote Sensing* **42**(7):1433–1443.

[12] Hollands , T. *Motion tracking of sea ice with SAR satellite data*. PhD thesis, 2012.

[13] Huang , Z., Shi , X., Zhang , C., Wang , Q., Cheung , K. C., Qin , H., Dai , J., & Li , H. (2022) Flowformer: A transformer architecture for optical flow. In *European conference on computer vision* pages 668–685. Springer.

[14] Hutchings , J., Bliss , A., Mahoney , A., Polashenski , C., Fedders , E., Fitzgereld , C., Manganini , K., Forster , K., & May , M. (2023) Sea ice drift tracks for the sea ice dynamic experiment (sidex) field campaign from global navigation satellite systems (gnss) ice trackers, alaska, 2021. *Arctic Data Center*

[15] Ilg , E., Mayer , N., Saikia , T., Keuper , M., Dosovitskiy , A., & Brox , T. (2017) Flownet 2.0: Evolution of optical flow estimation with deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* pages 2462–2470.

[16] Karvonen , J. (2012) Operational sar-based sea ice drift monitoring over the baltic sea. *Ocean Science* **8** (4):473–483.

[17] Lavergne , T., Eastwood , S., Teffah , Z., Schyberg , H., & Breivik , L.-A. (2010) Sea ice motion from low-resolution satellite sensors: An alternative method and its validation in the arctic. *Journal of Geophysical Research: Oceans* **115**(C10).

[18] Lei , Y., Gardner , A., & Agram , P. (2021) Autonomous repeat image feature tracking (autorift) and its application for tracking ice displacement. *Remote Sensing* **13**(4):749.

[19] Li , S. & Liu , W. (2022) Impacts of arctic sea ice loss on global ocean circulations and interbasin ocean heat exchanges. *Climate Dynamics* **59**(9):2701–2716.

[20] Malik , I. H., Ford , J. D., Clark , D. G., & Pearce , T. (2024) Climate change, mass casualty incidents, and emergency response in the arctic. *Environmental Research: Infrastructure and Sustainability* **4**(4):043002.

[21] Mehl , L., Schmalfuss , J., Jahedi , A., Nalivayko , Y., & Bruhn , A. (2023) Spring: A high-resolution high-detail dataset and benchmark for scene flow, optical flow and stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* pages 4981–4991.

[22] Meng , L., Yan , C., Lv , S., Sun , H., Xue , S., Li , Q., Zhou , L., Edwing , D., Edwing , K., Geng , X., & others (2024) Synthetic aperture radar for geosciences. *Reviews of Geophysics* **62**(3):e2023RG000821.

[23] Menze , M. & Geiger , A. (2015) Object scene flow for autonomous vehicles. In *Conference on Computer Vision and Pattern Recognition (CVPR)*

[24] Morimitsu , H. Pytorch lightning optical flow. 2021.

[25] Morimitsu , H., Zhu , X., Ji , X., & Yin , X.-C. (2024) Recurrent partial kernel network for efficient optical flow estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence 38*, pp. 4278–4286.

[26] Morimitsu , H., Zhu , X., Cesar , R. M., Ji , X., & Yin , X.-C. (2025) Dpflow: Adaptive optical flow estimation with a dual-pyramid framework. In *Proceedings of the Computer Vision and Pattern Recognition Conference* pages 17810–17820.

[27] Petrini , L., Cagnetta , F., Vanden-Eijnden , E., & Wyart , M. (2022) Learning sparse features can lead to overfitting in neural networks. *Advances in Neural Information Processing Systems* **35**:9403–9416.

[28] Roberts , A. F., Hunke , E. C., Kamal , S. M., Lipscomb , W. H., Horvat , C., & Maslowski , W. (2019) A variational method for sea ice ridging in earth system models. *Journal of Advances in Modeling Earth Systems* **11**(3):771–805.

[29] Sandven , S., Spreen , G., Heygster , G., Girard-Ardhuin , F., Farrell , S. L., Dierking , W., & Allard , R. A. (2023) Sea ice remote sensingrecent developments in methods and climate data sets. *Surveys in Geophysics* **44** (5):1653–1689.

[30] Scheuchl , B., Flett , D., Caves , R., & Cumming , I. (2004) Potential of radarsat-2 data for operational sea ice monitoring. *Canadian Journal of Remote Sensing* **30**(3):448–461.

[31] Shah , S. T. H. & Xuezhi , X. (2021) Traditional and modern strategies for optical flow: an investigation. *SN Applied Sciences* **3**(3):289.

[32] Steiner , N. S., Bowman , J., Campbell , K., Chierici , M., Eronen-Rasimus , E., Falardeau , M., Flores , H., Fransson , A., Herr , H., Insley , S. J., & others (2021) Climate change impacts on sea-ice ecosystems and associated ecosystem services. *Elem Sci Anth* **9**(1):00007.

[33] Sun , D., Yang , X., Liu , M.-Y., & Kautz , J. (2018) Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *Proceedings of the IEEE conference on computer vision and pattern recognition* pages 8934–8943.

[34] Teed , Z. & Deng , J. (2020) Raft: Recurrent all-pairs field transforms for optical flow. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16* pages 402–419. Springer.

[35] Vogel , C., Bauder , A., & Schindler , K. (2012) Optical flow for glacier motion estimation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **1**:359–364.

[36] Wang , X., Chen , R., Li , C., Chen , Z., Hui , F., & Cheng , X. (2022) An intercomparison of satellite derived arctic sea ice motion products. *Remote Sensing* **14**(5):1261.

[37] Wang , Y. & Deng , J. (2025) Waft: Warping-alone field transforms for optical flow. *arXiv preprint arXiv:2506.21526*

[38] Wang , Y., Lipson , L., & Deng , J. (2024) Sea-raft: Simple, efficient, accurate raft for optical flow. In *European Conference on Computer Vision* pages 36–54. Springer.

[39] Weickert , J., Bruhn , A., Brox , T., & Papenberg , N. A survey on variational optic flow methods for small displacements. *Mathematical models for registration and applications to medical imaging*, pages 103–136. 2006.

[40] Willmes , S., Heinemann , G., & Schnaase , F. (2023) Patterns of wintertime arctic sea-ice leads and their relation to winds and ocean currents. *The Cryosphere* **17**(8):3291–3308.

[41] Xian , Y., Petrou , Z. I., Tian , Y., & Meier , W. N. (2017) Super-resolved fine-scale sea ice motion tracking. *IEEE Transactions on Geoscience and Remote Sensing* **55**(10):5427–5439.

[42] Xu , H., Zhang , J., Cai , J., Rezatofighi , H., & Tao , D. (2022) Gmflow: Learning optical flow via global matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* pages 8121–8130.

[43] Xu , H., Zhang , J., Cai , J., Rezatofighi , H., Yu , F., Tao , D., & Geiger , A. (2023) Unifying flow, stereo and depth estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(11):13941–13958.

[44] Yastrebova , A., Höyhtyä , M., Boumard , S., Lohan , E. S., & Ometov , A. (2021) Positioning in the arctic region: State-of-the-art and future perspectives. *IEEE Access* **9**:53964–53978.

[45] Zhang , Z., Gupta , A., Jiang , H., & Singh , H. (2024) Neuflow v2: High-efficiency optical flow estimation on edge devices. *arXiv preprint arXiv:2408.10161*

[46] Zhao , S., Zhao , L., Zhang , Z., Zhou , E., & Metaxas , D. (2022) Global matching with overlapping attention for optical flow estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* pages 17592–17601.

[47] Zhao , Y. & Liu , A. K. (2001) Principal-component analysis of sea-ice motion from satellite data. *Annals of Glaciology* **33**:133–138.

[48] Zheng , Z., Nie , N., Ling , Z., Xiong , P., Liu , J., Wang , H., & Li , J. (2022) Dip: Deep inverse patchmatch for high-resolution optical flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* pages 8925–8934.

A   Appendix / supplemental material

A.1   Extended Background

A.1.1   Classical Methods for Sea Ice Motion Estimation

Traditional sea ice motion estimation techniques rely on block matching, cross-correlation, and feature tracking applied to satellite images [12]. Maximum cross-correlation (MCC) [17] and phase correlation [41] have been widely adopted due to their efficiency. Extensions include Laplacian pyramids [11] and wavelet-based methods [47], aimed at improving robustness. However, all remain limited in regions of thin ice, low texture, or under rapid deformation [22]. Sparse methods based on tie points [16] offer reliability in structured regions but fail to capture full ice dynamics, especially under divergence or ridging [28]. Dense optical flow offers a richer representation [35] but must overcome SAR-specific noise.

A.1.2   Deep Learning for Optical Flow

Deep learning introduced end-to-end frameworks for motion estimation, beginning with FlowNet [6] and FlowNet2 [15]. PWC-Net [33] introduced feature pyramids and cost volumes, while RAFT [34] advanced accuracy through recurrent refinement. More recently, transformer-based models such as Unimatch [43] and others exploit global feature matching and self-attention, capturing long-range dependencies in motion. These advances have shown strong performance in RGB benchmarks, but their transfer to SAR remains largely untested [22].

A.1.3 Applications in Climate Science and Navigation

Satellite-derived ice motion fields inform large-scale climate models of Arctic dynamics [29], while operational services (e.g., NIC, Copernicus Marine Service) use them to improve forecasts and navigation safety. Enhanced methods also support search and rescue, oil spill tracking, and polar field campaigns [20]. Improving motion estimation with deep learning could therefore benefit both scientific and operational communities.

A.2 Study area

Figure 3: Study area in the Beaufort Sea, Alaska, showing the region of interest used for benchmarking optical flow models.

Fig. 3 shows the study area in the Beaufort Sea, Alaska.

A.3 Dataset Details

The RADARSAT-2 ScanSAR dataset spans March 4–May 16, 2021, comprising 54 images at 20, 40, and 50 m spatial resolution. The dataset captures diverse Arctic sea ice states, including compact floes, divergence zones, and lead formation.

A.4 Preprocessing Pipeline

Images were orthorectified, co-registered using satellite ephemeris data, and normalized to mitigate SAR-specific intensity variations.

Results Across All Models

While the main paper highlights the top five models, Table 2 reports full results for all 48 evaluated architectures. This provides a comprehensive reference for future studies comparing deep optical flow methods in geophysical contexts.

Reference Metrics from the PTLFlow Library

Table 3 compiles the publicly available evaluation metrics (EPE and Fl-all) from the PTLFlow library [24], covering a wide range of optical flow architectures. These values correspond to results on standard computer vision benchmarks and are included here to provide a reference context for the models evaluated in this study. The metrics in this table are not directly comparable to those reported for the RADARSAT-2 benchmark in this study, due to differences in data modality (RGB vs. SAR), ground-truth density, and spatial scale of motion.

Table 2: Endpoint Error (EPE) mean and standard deviation, and Fl-all mean and standard deviation for all 48 evaluated models.

Model Name	EPE Mean [px]	EPE Std [px]	Fl-all Mean [%]	Fl-all Std [%]
dip_kitti [48]	8.04	8.98	51.3	33.67
dip_sintel [48]	6.29	5.09	42.52	30.06
dip_things [48]	60.15	127.68	49.24	34.36
dpflow_chairs [26]	6.27	5.3	43.08	29.83
dpflow_kitti [26]	8.55	9.41	63.13	34.72
dpflow_sintel [26]	7.76	6.88	53.07	37.25
dpflow_things [26]	7.67	11.55	43.17	30.75
flowformer_chairs [13]	7.19	7.98	44.92	31.13
gmflow_chairs [42]	8.35	11.63	55.88	36.24
gmflow_kitti [42]	7.21	6.48	55.74	31.57
gmflow_refine_chairs [42]	7.57	10.66	47	31.92
gmflow_refine_kitti [42]	6.85	6.09	54.38	30.33
gmflow_refine_sintel [42]	6.12	5.36	44.9	30.53
gmflow_refine_things [42]	6.26	5.27	44.71	29.4
gmflow_sintel [42]	7.62	9.57	54.05	33.84
gmflow_things [42]	9.02	18.45	52.38	33.89
gmflownet_kitti [46]	8.04	7.15	53.43	35.19
gmflownet_mix_sintel [46]	7.23	7.53	45.98	32.27
gmflownet_mix_things [46]	7.96	13.29	44.47	31.09
neuflow2_mixed [45]	7.43	8.6	47.41	32.5
pwcnet_things [33]	8.54	16.5	44.76	28.94
raft_kitti [34]	9.25	11.37	55.86	34.28
raft_sintel [34]	6.88	7.04	44	31.15
raft_things [34]	6.88	6.25	45.37	31.71
rpknet_chairs [38]	7.73	12.01	43.3	30.56
rpknet_kitti [38]	6.26	5.32	44.41	30.79
rpknet_sintel [38]	7.51	11.33	42.89	30.34
rpknet_things [38]	7.02	7.96	43.43	30.82
sea_raft_l_chairs [38]	17.15	19.99	47.55	30.57
sea_raft_l_kitti [38]	17.81	13.6	86.19	24.53
sea_raft_l_sintel [38]	8.51	12.75	45.89	33.02
sea_raft_l_spring [38]	7.5	10.66	43.18	30.39
sea_raft_l_tartan [38]	7.84	11.54	47.37	32.93
sea_raft_l_things [38]	126.52	169.42	61.53	39.04
sea_raft_m_chairs [38]	8.85	10.16	45.33	30.92
sea_raft_m_kitti [38]	12.2	12.33	82.35	27.84
sea_raft_m_sintel [38]	8.3	12.26	46.72	32.89
sea_raft_m_spring [38]	7.56	11.09	43.08	30.34
sea_raft_m_tartan [38]	7.52	10.91	45.32	31.01
sea_raft_m_things [38]	18.11	36.08	47.89	34.11
sea_raft_s_chairs [38]	7.97	11.16	46.59	32.31
sea_raft_s_kitti [38]	12.4	13.71	63.83	34.47
sea_raft_s_sintel [38]	9.36	10.87	68.44	28.93
sea_raft_s_spring [38]	7.75	12.26	43.12	30.34
sea_raft_s_things [38]	7.96	11.29	44.66	32.03
unimatch_things [43]	9.02	18.45	52.38	33.89
waft_chairs [37]	8.87	10.75	52.92	35.54
waft_things [37]	10.03	25.67	49.59	34.45

Table 3: Summary of end-point error (EPE) and flow outlier rate (Fl-all) metrics available in the PTLFlow library [24]. Only metrics officially reported by the library are included. These results provide a reference for the relative performance of standard optical flow models under common computer vision benchmarks.

model	checkpoint	sintel-clean-epe	sintel-clean-flall	sintel-final-epe	sintel-final-flall	kitti-2012-epe	kitti-2012-flall	kitti-2015-epe	kitti-2015-flall
dip	kitti	1.176	3.164	2.168	6.721	0.945	2.861	1.129	3.424
dip	sintel	0.504	2.037	1.494	5.506	1.619	5.922	3.977	11.139
dip	things	1.332	3.414	3.187	7.939	1.74	6.918	4.426	13.349
flowformer	chairs	1.952	6.208	4.156	10.687	4.217	23.041	9.182	33.215
gmflow	chairs	3.225	15.873	4.432	18.904	8.623	47.849	16.542	54.489
gmflow	kitti	2.858	11.174	3.906	14.48	2.077	9.265	2.104	7.889
gmflow_refine	chairs	2.305	7.214	3.637	10.668	5.636	23.869	13.161	37.014
gmflow_refine	sintel	0.762	2.414	1.11	4.279	1.638	6.557	1.975	7.728
gmflow_refine	things	1.084	3.652	2.474	7.331	3.132	13.533	6.747	21.673
gmflow_refine	kitti	2.443	8.395	3.645	12.582	1.59	6.479	1.363	5.119
gmflownet	things	1.309	3.908	2.727	7.827	1.963	8.418	4.658	15.837
gmflownet	kitti	4.369	10.123	6.239	14.357	1.207	4.001	0.621	1.502
neuflow2	mixed	0.999	3.733	1.661	6.513	1.293	4.691	1.666	6.106
pwcnet	things	2.643	8.811	4.059	12.592	4.042	20.243	10.288	33.014
raft	things	1.458	4.173	2.684	7.721	2.135	9.257	5.018	17.046
raft	sintel	0.763	2.855	1.218	4.936	1.338	4.587	1.555	5.573
raft	kitti	4.544	10.511	6.157	14.794	1.283	4.272	0.643	1.58
rpknet	chairs	2.406	7.341	3.588	10.697	3.965	22.812	10.878	33.835
rpknet	kitti	1.478	4.605	2.37	7.86	1.171	3.817	0.946	2.904
rpknet	sintel	0.552	2.19	0.813	3.659	1.192	4.112	1.257	4.273
rpknet	things	1.12	3.516	2.449	7.08	1.674	6.817	3.783	12.998
sea_raft_l	spring	1.197	3.074	2.207	6.131	2.342	7.856	4.371	12.424
sea_raft_l	kitti	0.617	2.025	0.931	3.324	1.049	3.219	0.817	2.267
sea_raft_l	things	1.226	3.116	3.444	7.053	2.061	7.898	3.723	12.377
sea_raft_l	chairs	2.441	5.422	9.272	11.951	5.292	31.121	10.5	37.727
sea_raft_l	tartan	2.316	7.701	5.549	15.499	1.276	4.205	3.306	9.977
sea_raft_l	sintel	0.444	1.597	0.613	2.292	1.086	3.391	0.892	2.571
sea_raft_m	tartan	2.366	6.752	5.203	14.289	1.385	4.467	3.568	9.887
sea_raft_m	chairs	2.27	5.382	5.744	11.259	4.359	25.483	9.159	33.925
sea_raft_m	things	1.272	3.235	3.853	7.343	1.971	7.406	4.293	13.935
sea_raft_m	sintel	0.434	1.573	0.584	2.197	1.125	3.538	0.88	2.551
sea_raft_m	kitti	0.646	2.136	0.958	3.49	1.078	3.373	0.808	2.213
sea_raft_m	spring	1.389	3.436	2.3	6.455	2.869	8.532	5.812	14.49
sea_raft_s	chairs	2.052	5.406	4.711	10.625	4.401	27.262	9.579	36.224
sea_raft_s	things	1.283	3.34	3.729	7.307	2.001	8.172	4.433	14.694
sea_raft_s	sintel	0.547	1.92	0.783	3.009	1.153	3.754	1.054	3.151
sea_raft_s	kitti	1.067	3.072	1.756	5.553	1.093	3.589	0.935	2.7
sea_raft_s	spring	1.697	4.064	2.573	7.095	3.552	10.453	7.906	18.163