

River Ice Fine-grained Segmentation: A GF-2 Satellite Image Dataset and Deep Learning Benchmark

Chenxu Wei, Haoxuan Li, Liang Chen, Haohao Zhou, Omirzhan Taukebayev, Wencong Wu, Amirkhan Temirbayev, Lin Han, Lingyan Ran, Hanlin Yin, Peng Wang, Junrui Liu, Xiuwei Zhang, Yanning Zhang,
Fellow IEEE

Abstract—Semantic segmentation of river ice image serves as a critical technological foundation for hydrological monitoring and ice flood early warning system. Current publicly available river ice datasets predominantly utilize UAV-captured image and ground-based photographic observations. To address the limitations of spatial coverage in existing datasets, we present NWPU_YRCC_GFICE - a satellite remote sensing dataset constructed from multi-spectral GF-2 satellite images. The dataset innovatively categorizes river ice into six fine-grained classes across freeze-thaw cycles and covers river ice data from Yellow River (Ningxia-Inner Mongolia section) spanning the past 10 years. We further establish a comprehensive deep learning benchmark, which evaluates 33 state-of-the-art segmentation models and two improved segmentation models based on YOLO and Segformer architecture, separately. Experiments are conducted on the NWPU_YRCC_GFICE dataset and three public river ice datasets (NWPU_YRCC_EX, NWPU_YRCC2, and Alberta river ice segmentation dataset). The proposed models exhibit excellent performance, surpassing the state-of-the-art methods. The presented NWPU_YRCC_GFICE dataset and benchmark enriches the river ice dataset and favors in promoting fine-grained river ice segmentation research from satellite view. Our dataset and code is available at https://github.com/ASGOLabMultisourceCooperationGroup/NWPU_YRCC_GFICE

Index Terms—River ice dataset, Fine-grained semantic seg-

mentation, YOLO, Segformer

I. INTRODUCTION

River ice represents a dynamic natural phenomenon with significant impacts on both ecological systems and human societies. In regions of the Northern Hemisphere where temperatures drop below 0°C, river ice forms as water freezes under cold weather conditions [1]. The accumulation of river ice can provoke flooding via ice jams, hinder navigation, and disturb local ecosystems. Due to these potential threats, efficient monitoring of river ice is imperative for ensuring public safety, anticipating disasters, and enhancing resource management [2], [3]. A critical technological foundation of river ice monitoring is river ice segmentation, which involves classifying different elements such as ice, water, and riverbanks at the pixel-level in image data. It further aids in distinguishing the various states of river ice. The accurate segmentation is vital for a precise analysis of river ice dynamics, thereby advancing monitoring and management practices.

Semantic segmentation has made significant progress since the advent of Fully Convolutional Networks (FCN) [4], which laid the foundation for modern segmentation techniques. Building on the FCN framework, models such as DeepLab [5]–[8], U-Net [9], U-Net++ [10], HRNet [11], and RefineNet [12] significantly advanced the field. Transformer-based models, such as SegFormer [13] and the Segmentation Transformer (SETR) [14], brought innovation through powerful global context modeling. Models like BiseNet [15], PIDNet [16], and YOLOv8 [17] have focused on balancing accuracy with processing speed, making them suitable for real-time applications. Additionally, large-scale models like the Segment Anything Model (SAM) [18] and others [19], [20], [21] have enhanced scalability and efficiency, enabling diverse applications.

With the continuous development of semantic segmentation technology, various models have been applied to river ice segmentation. For instance, IceHrNet [22], based on HRNet [11], has been used for high-resolution river ice analysis, while ICENet [23], built on the BiseNet [15] framework, targets river ice segmentation. Singh et al. [24] applied CNN-based models like U-Net [25], SegNet [26], DeepLab [6], and DenseNet [27] to classify river ice into categories such as water, drift ice, and anchor ice. UAV-based datasets, such as the Alberta

This work is supported in part by the National Key R&D Program of China (2024YFE0213600), the National Natural Science Foundation of China (62476226, 61971356), Natural Science Basic Research Program of Shaanxi (2024JC-DXWT-07 and 2024JC-YBQN-0719), National Key Research, Guangdong Basic and Applied Basic Research Foundation (2024A1515030186), Natural Science Foundation of NingBo (2023J262). (Corresponding author: Xiuwei Zhang, Junrui Liu and Hanlin Yin)

Chenxu Wei, Haoxuan Li, Wencong Wu, Lingyan Ran, Hanlin Yin; Peng Wang, Junrui Liu, Xiuwei Zhang and Yanning Zhang are with the School of Computer Science, Shaanxi Provincial Key Laboratory of Speech and Image Information Processing, and the National Engineering Laboratory for Integrated Aerospace-GroundOcean Big Data Application Technology, the Ningbo Institute of Northwestern Polytechnical University, Northwestern Polytechnical University, Xi'an 710072, China (email: weichenxu@mail.nwpu.edu.cn; li_haoxuan@mail.nwpu.edu.cn; wencongwu@mail.nwpu.edu.cn; lran@nwpu.edu.cn; iverlon1987@126.com; peng.wang@nwpu.edu.cn; liu.junrui@nwpu.edu.cn; xwzhang@nwpu.edu.cn; ynzhang@nwpu.edu.cn).

Omirkhan Taukebayev and Amirkhan Temirbayev is with Al-Farabi Kazakh National University, Republic of Kazakhstan (email: omirzhan.taukebayev@kaznu.edu.kz; amirkhan.temirbayev@kaznu.edu.kz).

Liang Chen, Haohao Zhou are with the Information Center of Yellow River Conservancy Commission, ZhengZhou 450004, China (email:chlq564@163.com; 2901151727@qq.com).

Lin Han is with the School of Geography and Tourism, Zhengzhou Normal University, ZhengZhou 450044, China (email:hanlin@zznu.edu.cn).

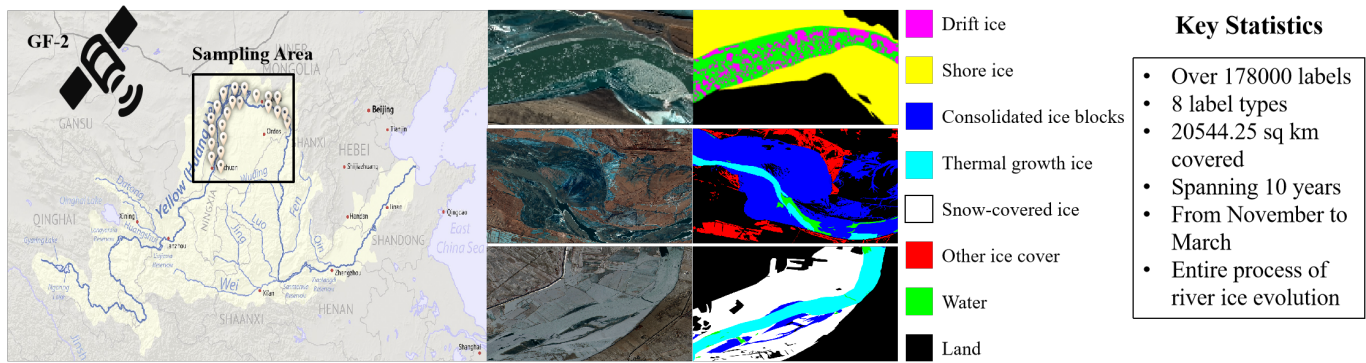


Fig. 1: NWPU_YRCC_GFICE represents an extensive remote sensing dataset obtained via the GF-2 satellite and annotated by hand. The dataset focuses on the redundant classification and segmentation of river ice. It categorizes river ice into six types: drift ice (pink), shore ice (yellow), thermal growth ice (dark blue), consolidated ice blocks (light blue), mixture of snow with ice or others (white) and other ice cover (red). Additionally, it includes water (green) and land (black).

River Ice Segmentation Dataset [24], have played a pivotal role in advancing segmentation, with models like ICENet [23], ICENetV2 [28], and FastICENet [29] demonstrating success in segmenting ice and water features. Similarly, fixed camera datasets [30], [31] are also widely used for river ice monitoring.

In addition to UAV and camera-based datasets, satellite data has been extensively employed for river ice monitoring. For example, Temimi et al. [32] use NOAA-20 and NPP satellites for near-real-time monitoring of ice in northern U.S. and Canadian basins. Muhammad et al. [33] apply MODIS data from Terra and Aqua satellites to monitor freeze-thaw conditions on the Mackenzie River, while B. Altena et al. [2] combine Sentinel-2 and PROBA-V data for ice movement estimation along the Lena River. Additionally, [34] uses Sentinel-1 for detecting ice jams in Finnish rivers. While these studies focused on large-scale tasks like freeze-thaw detection or assessing the extent of ice cover, fine-grained classification of river ice remains essential. Different types of river ice, including drift ice, shore ice, and flat ice, provide key insights into the river's freezing conditions and help predict ice jam formation, critical for flood forecasting. The absence of large-scale, high-quality annotated datasets for river ice segmentation in satellite image remains a critical bottleneck in advancing river ice monitoring research.

To explore detailed river ice classification, we select Ningxia-Inner Mongolia section of Yellow River as our study area, in which river ice phenomenon is very typical. Compared with UAV-captured image and ground-based photographic observation, satellites provide a larger view. We gather GF-2 satellite images of the studied area spanning ten years and covering the freeze-up to break-up period from November to March, and construct a satellite remote sensing dataset, named NWPU_YRCC_GFICE, for fine-grained river ice segmentation. It categorizes river ice into six fine-grained classes across freeze-thaw cycle. Moreover, we develop a comprehensive deep learning benchmark, which evaluates 35 segmentation models based on NWPU_YRCC_GFICE, including two improved segmentation models based on YOLO and Segformer architecture, separately. The presented NWPU_YRCC_GFICE

dataset and benchmark favors in promoting fine-grained river ice segmentation research in a large-scale view. The main contributions are summarized as follows:

- To advance fine-grained river ice segmentation research, we present a satellite remote sensing dataset, namely NWPU_YRCC_GFICE. All images are captured by GF-2 satellite in Ningxia-Inner Mongolia section of Yellow River, in which river ice phenomenon is very typical. In total, it contains 28,378 images with a size of 768×768 pixels. Pixel-level annotations are performed on the NWPU_YRCC_GFICE dataset, including 8 categories covering six fine-grained types of river ice, land, and water. To date, this dataset is the first publicly available satellite remote sensing dataset for fine-grained river ice classification.
- Based on NWPU_YRCC_GFICE, we establish a comprehensive deep learning benchmark for fine-grained river ice segmentation, evaluating 33 state-of-the-art segmentation models and providing two recommended solutions: ICEYOLO and ICEFormer. This benchmark offers systematic performance comparison and delivers practical recommendations for different application scenarios. ICEYOLO is based on YOLOv8 to take advantage of YOLO's real-time speed and lightweight architecture, providing an optimal solution for real-time monitoring applications. To adapt irregular shape characteristics of river ice, two improvements are adopted, i.e., adding attention at different scale in encoding stage and designing an adaptive keypoint redistribution algorithm in training stage. ICEFormer is constructed on Segformer architecture and extended by a multiscale fusion, offering the highest accuracy solution for precision-critical applications.
- Experiments are conducted on the NWPU_YRCC_GFICE dataset and three public river ice datasets (NWPU_YRCC_EX, NWPU_YRCC2, and Alberta river ice segmentation dataset). The proposed models exhibit excellent performance, surpassing the state-of-the-art methods.

II. RELATED WORK

A. Semantic Segmentation

Since the advent of fully convolutional networks (FCNs) [4], which eliminates the need for traditional patch-based methods and facilitates the application of deep learning techniques to segmentation task, the field of semantic segmentation has made great progress and promotion. Following the FCN, several excellent models emerge. The DeepLab series [5]–[7] introduce atrous (dilated) convolution, which expands the receptive field without losing resolution, improving feature extraction from larger contextual regions. U-Net [9] and its evolution U-Net++ [10] advance segmentation by using skip connections and nested skip pathways, which enhance feature propagation and segmentation precision, especially in biomedical and high-resolution imaging. RefineNet [12] focuses on refining coarse predictions through networks designed for multi-path refinement, enhancing the extraction of fine-grained features. HRNet [11] distinguishes itself by maintaining high-resolution representations throughout the entire network, thereby setting a high standard for preserving spatial information across various segmentation applications. Collectively, these models, with their innovative architectures, have driven the field forward by enhancing feature extraction, contextual understanding, and precise prediction, paving the way for more sophisticated semantic segmentation methods.

With the emergence of Transformer and its outstanding performance in various fields, recent innovations in semantic segmentation have led to the development of Transformer-based models that emphasize modeling global context. Initially designed for image classification, the Vision Transformer (ViT) [35] has been successfully adapted for segmentation by utilizing a self-attention mechanism to capture broad dependencies. This method improves segmentation accuracy by thoroughly understanding the global context. Building on this, the Segmentation Transformer (SETR) [14] refines the transformer framework specifically for pixel-level segmentation by replacing typical convolutions with full self-attention, achieving impressive results on complex datasets. More recently, the Swin Transformer [36] has been acknowledged for its use of a shifted window approach in self-attention. This technique effectively balances the extraction of both local and global features while reducing computational demands. In addition, SegFormer [13] introduces an efficient hierarchical transformer structure that boosts scalability and accuracy, advancing the role of transformers in semantic segmentation. Further, MMT (Mixed-Mask Transformer) [37] introduces a mixed-mask attention mechanism and progressive multiscale learning strategy to address the foreground–background imbalance problem in high-resolution remote sensing scenes, achieving state-of-the-art performance on standard benchmarks such as ISPRS Potsdam and Vaihingen.

Simultaneously, there has been an increasing emphasis on balancing segmentation accuracy and computational efficiency, which is especially important for real-time applications requiring fast processing. Models such as BiSeNet [15], PIDNet [16], and YOLOv8 [17] employ lightweight architectures to decrease processing time with negligible losses in accuracy,

rendering them ideal for real-time use in resource-constrained environments. In addition, Cross Fusion Net (CF-Net) [38] is a lightweight segmentation network that efficiently captures small-scale semantic details through attention-based fusion. These models achieve remarkable segmentation through techniques such as enhanced feature extraction via parallel convolutional pathways and the integration of effective attention mechanisms. Furthermore, larger models like the Segment Anything Model (SAM) [18], along with associated frameworks [19], [20], [21], have been designed to prioritize scalability, robustness, and efficiency. [39] proposes a postprocessing framework that directly integrates SAM's raw outputs with the predictions of semantic segmentation models, effectively enhancing mask quality without requiring additional training. These models adeptly handle a wide range of segmentation tasks, from fine-grained pixel-level delineation to comprehensive image processing, delivering excellent outcomes across numerous applications. Their ability to scale and adapt to various environments has become a crucial advancement in semantic segmentation, enabling high-quality segmentation on a larger scale with faster processing speeds.

B. River Ice Segmentation

With the development of remote sensing technologies, satellites, UAVs, and fixed cameras have been adopted to monitor the formation and development of river ice from different views.

Before deep learning, river ice segmentation relies on manual feature extraction, rule-based methods, and traditional machine learning techniques. SAR image is commonly used, with techniques such as k-means clustering and grouping pixels based on grayscale or texture characteristics, but this method struggles with unbalanced data distributions [40]. To detect river ice breakup dates, Beaton et al. [41] propose a calibrated threshold method to distinguish open water and not open water (solid ice, cloud or a combination of the two) and mitigate the impact of cloud interference. Multithreshold techniques [42] are also used for pixel classification, but are still vulnerable to environmental noise [43]. Bharathi et al. [44] introduce a texture-based color segmentation method for infrared river ice images, using K-means clustering combined with Gabor filters for texture segmentation. These methods mainly focus on river ice extraction, while their accuracy is limited by noise, lighting, and the complexity of manual feature extraction.

At this stage, satellite data play a crucial role in advancing river ice monitoring. Most approaches adopt satellite images to perform a global analysis of river ice distribution. Data from different sources, such as VIIRS and MODIS, or TanDEM-X and Landsat [45], [46], are applied to improve the segmentation accuracy. MODIS data are used to follow freeze-thaw cycles in rivers such as the Mackenzie River [33]. These studies have largely concentrated on large-scale monitoring activities, such as detecting freeze-thaw cycles and gauging ice cover extents. In addition to satellite observations, ground-based sensing techniques have also been applied to river ice monitoring. Purnell et al. [47] employ GNSS-IR in

combination with SAR image and machine learning to capture river ice breakup processes with high temporal sensitivity.

The rapid development of deep learning has significantly advanced intelligent river ice monitoring. Singh et al. [24] introduce CNN-based segmentation models (e.g., U-Net [9], SegNet [26]) to classify river ice into water, drift ice, and anchor ice, achieving pixel-level accuracy. Subsequent studies further optimize feature representation for ice segmentation: ICENet [23] fuses positional and channel-wise attention to enhance ice boundary delineation, while ICENetv2 [28] incorporates multiscale fusion to address fine-grained ice variations in UAV-captured image. FastICENet [29] introduces a lightweight architecture with parallel convolutional pathways, achieving real-time inference speeds while maintaining high segmentation accuracy for dynamic ice monitoring. Ansari et al. [48] propose IceMaskNet, an instance segmentation model based on Mask R-CNN with a ResNetV2-50 backbone, aimed at improving segmentation efficiency and suitability for UAV-based river ice monitoring. Fu et al. [49] propose CSEU-Net, a novel U-Net variant integrating a ConvNeXt-U backbone and SE attention modules, specifically designed for segmenting river ice floes in UAV-based grayscale image. Zhao et al. [50] develop a transfer learning-based framework integrating channel-spatial attention mechanisms and pyramid pooling to address fine-grained river ice segmentation in high-latitude urban environments. The above algorithms utilize UAV images to segment river ice. Promoted by the fast development of UAV technologies, UAVs with these excellent algorithms provide a fast, convenient, wide-range, and high-resolution way to monitor river ice.

In parallel with UAV-based efforts, fixed ground-based cameras have also been adopted for long-term river ice monitoring using deep learning techniques. Ansari et al. [51] apply a Mask R-CNN-based instance segmentation algorithm to oblique shore-based image captured along the Dauphin River, enabling effective segmentation and classification of river ice types from camera-collected images. RIce-Net [52] is presented to segment images from the U.S. Geological Survey fixed river camera network to calculate the fraction of ice coverage, to automatically generate ice flags. It contains a two-stage process: introducing a binary classifier based on FCN with Softmax activation function before the segmentation step to filter out ice-free images and then inputting the features of the ice-affected images into a pyramid attention network (PAN) to segment the ice on the water surface. Pei et al. [30] propose a hybrid pipeline for the North Saskatchewan River ice analysis, including classification, geometric rectification, and segmentation, in which a simple CNN architecture is adopted to classify the image into four classes and a UNet architecture is utilized to segment river ice. These works demonstrate the feasibility and robustness of applying deep learning to fixed-camera image, particularly for long-term and oblique-view monitoring of river ice.

Satellite image has also been increasingly integrated with deep learning methods. Temimi et al. [32] apply the U-Net architecture to classify NOAA-20 and NPP satellite images into seven types, including water, land, vegetation, snow, river ice, cloud, and cloud shadow. Altena et al. [2] use Sentinel-2

and PROBA-V to estimate ice block motion, while Sentinel-1 is utilized for detecting ice jams under cloudy conditions [34]. These methods focus on large-scale dynamic monitoring but lack the granularity required for fine-level classification of specific ice types.

Although significant progress has been made in river ice monitoring, fine-grained river ice segmentation across freeze-thaw cycles remains a critical challenge. This task is especially important for accurately monitoring ice dynamics and forecasting ice jams.

III. NWPU_YRCC_GFICE DATASET

A. Motivation

High quality river ice datasets play an indispensable role in river ice monitoring based on deep learning. As shown in Table I, the existing public river ice datasets can be folded into three groups according to the data collection platform, namely UAV-based datasets, satellite-based datasets and ground-camera-based datasets.

Several UAV-based datasets have been developed to advance the intelligent monitoring of river ice. NWPU_YRCC dataset [23], collected from UAV image over the Ningxia-Inner Mongolia reach of China's Yellow River, includes 814 images annotated into three categories (ice, water, and shore) with a resolution of 1600×640 pixels. NWPU_YRCC_EX dataset [29], an extension of the previous dataset, consists of 887 images, also from the Yellow River, and covers three categories. Similarly, NWPU_YRCC2 dataset [29] expands to 1,525 images with four categories (drift ice, shore ice, water, and shore), addressing finer-grained classification needs for ice management. Singh et al. [24] propose a comprehensive benchmark on surface ice classification, leveraging high-resolution images and videos captured by UAVs and bridge-mounted cameras over Alberta rivers. SHR_RIDS dataset [50], comprising 1,161 UAV images captured from the middle reaches of the Songhua River near Harbin, China, focuses on high-latitude urban river ice with three pixel-level classes: ice cover, drift ice, and water. While the dataset collected from the Dauphin River, presented by Ansari et al. [51], includes 115 images and six categories (frazil pan, broken ice, frazil slush, border ice, ice cover, and open water). The dataset presented by Fu et al. [49] focuses on the Yellow River's Binzhou section. It comprises 65 original UAV images augmented to 27,180 training samples through Gaussian noise, blurring, and rotation, targeting binary segmentation (ice vs. water). In addition, the dataset introduced [53] by comprises 1,101 UAV-captured images from the Songhua River in Northeast China, covering ice channels formed by air-cushion icebreakers. This dataset provides pixel-level annotations for three categories: ice sheet, ice channel, and background. These datasets, based on UAV-captured image, are invaluable for understanding river ice behavior.

In addition to UAV-based datasets, several ground fixed-camera-based datasets have been introduced. The IPC_RI_IDS dataset [31], comprising 15,600 fixed-camera images captured from the Nenjiang River in China, provides continuous monitoring of ice dynamics across five distinct freeze-thaw stages

TABLE I: A collection of river ice segmentation datasets categorized by collection platform.

#	Dataset	Collection Platform	Volume	Classes	Location	Resolution
1	NWPU_YRCC [23]	UAV	814	3	Yellow River, China	$1,600 \times 640$
2	NWPU_YRCC_EX [29]	UAV	887	3	Yellow River, China	$1,600 \times 640$
3	NWPU_YRCC2 [29]	UAV	1,525	4	Yellow River, China	$1,600 \times 640$
4	Alberta River Ice [24]	UAV+Ground	255	3	Rivers across Alberta, Canada	$1,134 \times 1,009 \sim 1,462 \times 1,463$
5	SHR_RIDS [50]	UAV	1,161	3	Songhuajiang River, China	$3,840 \times 2,160$
6	Ansari et al. [48]	UAV	115	6	Dauphin River, Canada	$1,280 \times 720 \sim 5,472 \times 3,648$
7	Fu et al. [49]	UAV	65	2	Yellow River, China	$1,920 \times 1,080$
8	Zhao et al. [53]	UAV	1,101	3	Songhuajiang River, China	$3,840 \times 2,160$
9	Ansari et al. [51]	Ground	1,795	5	Dauphin River, Canada	—
10	IPC_RI_IDS [31]	Ground	15,600	5	Nenjiang River, China	$1,280 \times 720$
11	Pei et al. [30]	Ground	1,893	4	N. Saskatchewan River, Canada	820×500
12	NIMS [52]	Ground	1,829	3	Milwaukee River, U.S.	$1,600 \times 640 \sim 1,281 \times 1,081$
13	NWPU_YRCC_GFICE	Satellite	28,378	8	Yellow River, China	768×768

(ice frozen, break-up initiation, drifting, break-up completion, and ice-free), with pixel-level annotations for ice, water, and background categories. Furthermore, the dataset presented by Ansari et al. [48] includes 1,795 oblique shore-based images captured from the Dauphin River in Canada, annotated into five categories: open water, border ice, ice cover, frazil slush, and frazil pan/surface collar ice. The dataset introduced by Pei et al. [30] consists of long-term time-lapse images captured by a public rooftop camera with a distant, oblique view of the North Saskatchewan River in Canada. It contains 1,893 images and spans multiple freeze-up and breakup periods across several years, with manual annotations of ice pans and full ice cover used for segmentation and shape analysis. The NIMS images dataset [52], captured using ground-based cameras along the Milwaukee River in Wisconsin, U.S., is annotated for river ice segmentation into three categories: water, ice, and other. While the ground mounted fixed-camera approach allows for continuous monitoring, it has limitations in terms of spatial coverage.

Satellite image addresses the limitations of UAVs and stationary cameras by providing extensive spatial coverage, enabling the observation of expansive river ice dynamics. Upon examining existing datasets for semantic segmentation in river ice monitoring, despite there are many UAV and ground camera-based datasets, no publicly satellite remote sensing image dataset for river ice classification and fine-grained semantic segmentation is available. To advance deep learning research in this area, we have created a multi-spectral satellite image dataset specifically for river ice fine-grained segmentation. This dataset is a collaborative effort between Northwestern Polytechnical University and the Yellow River Conservancy Commission Information Center, and it is designated as the NWPU_YRCC_GFICE River Ice Dataset. Here, “NWPU_YRCC” signifies Northwestern Polytechnical University/Yellow River Conservancy Commission, while “GFICE” denotes river ice images captured by the GF-2 satellite. It comprises 28,378 satellite images from the Yellow River in China, spanning multiple freeze-thaw cycles. Totally, it covers eight categories and classifies river ice into six distinct types with fine pixel-level manual annotation.

Our motivation of NWPU_YRCC_GFICE River Ice Dataset is to advance satellite-based river ice segmentation and make more accurate, fine-grained and large-scale monitoring of river ice dynamics enabled.

B. Dataset Constructions

Data collection. To automatically monitor river ice dynamics throughout the ice season in a large-scale view, we collect 37 GF-2 satellite multi-spectral images from 2015 to 2024, covering the months from November to March, including periods of ice formation, growth, and breakup. These images are acquired from the Ningxia-Inner Mongolia section of the Yellow River. The longitude of the covered area is 106.3–111.5, and the latitude is 38.3–40.9. As shown in Figure 1, it covers the entire “U”-shaped part of the river. We choose this area because the ice phenomenon in this area is the most typical in the Yellow River Basin. The GF-2 satellite data consists of two parts: a four spectral (red, green, blue, and near infrared) multi-spectral image with a resolution of 4 meters and a panchromatic image with a resolution of 0.8 meters. The size of these images ranges from $46,320 \times 33,169$ pixels to $39,392 \times 30,008$ pixels.

Data preprocessing. Some preprocessing procedures are required to perform on original satellite images, including panchromatic image orthorectification, multi-spectral and panchromatic image matching, multi-spectral image orthorectification, and pan-sharpening of orthorectified panchromatic and multi-spectral image. After preprocessing, we obtain orthorectified multi-spectral images with a resolution of 0.8 meters, which are as the input to manual annotation.

Manual annotation. There are several tools for semantic segmentation annotations, such as Image Labeler and Labelme, which are typically designed to annotate objects with simple shapes, like vehicles. Practical annotation experience revealed challenges with these tools when dealing with satellite images of river ice, which exhibit substantial variability in size and form. Therefore, we adapt ENVI Classic 5.3 software to produce three-channel pseudo-RGB images. Then, these images are finely annotated in Photoshop, distinguishing eight categories: shore ice, drift ice, consolidated ice blocks, thermal

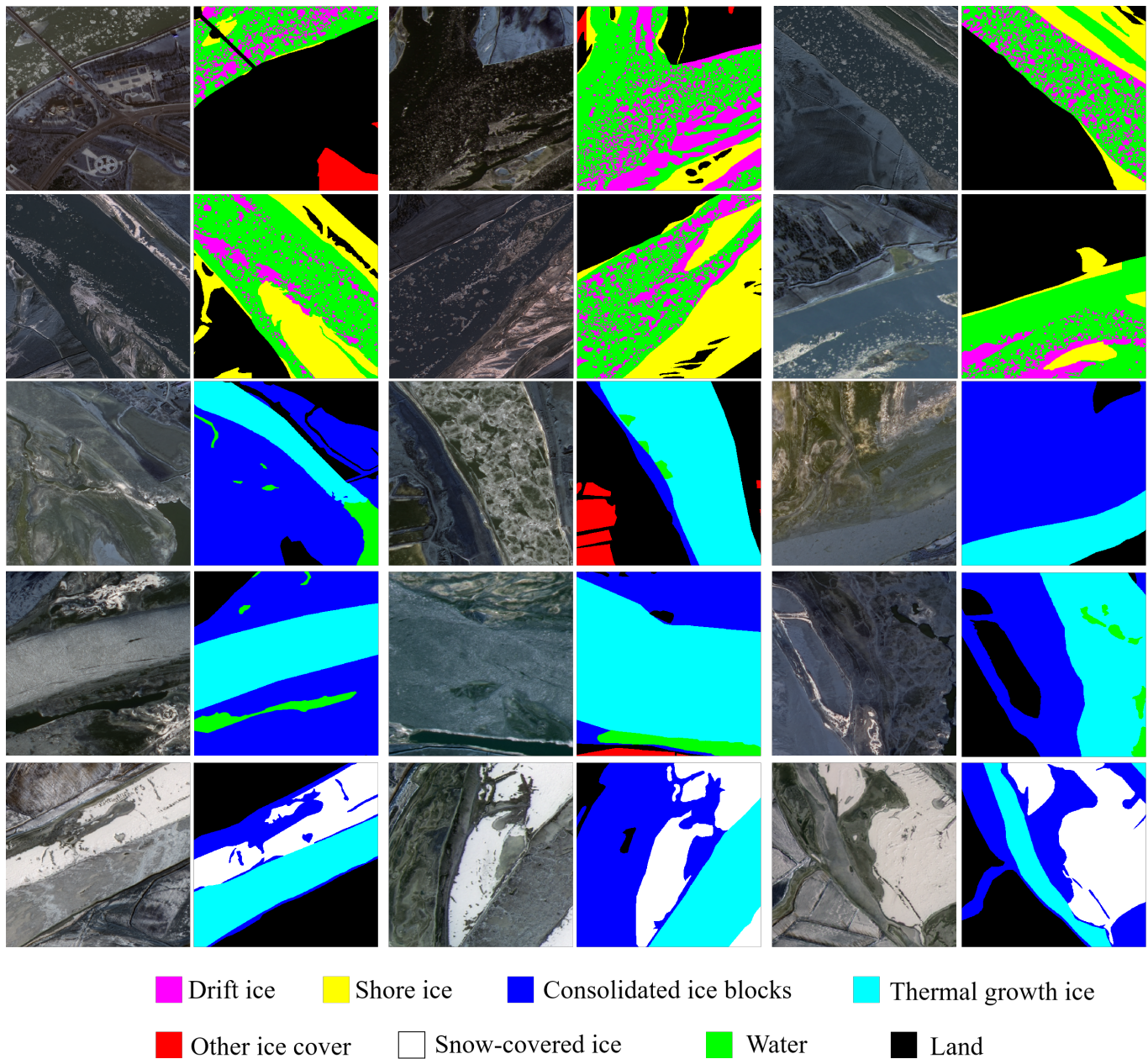


Fig. 2: Visualization of River Ice Segmentation across Various Ice Types. This figure presents a set of images from the NWPU_YRCC_GFICE dataset, showcasing the segmentation of different river ice categories. Each pair consists of the original image on the left and its corresponding segmentation mask on the right. The categories include Drift Ice, Bank Ice, Consolidated Ice Blocks, Thermal Growth Ice, Other Ice Cover, Mixture of Snow with Ice or Others, Water, and Other. The segmentation process is highlighted by distinct color coding for each ice type, providing a clear illustration of the dataset's detailed classification.

growth ice, mixture of snow with ice or others, water, other ice cover, and land. This process is highly labor-intensive, and ultimately complete the complete annotation of 37 images.

Dataset description. Since the size of each satellite image is approximately $40,000 \times 30,000$ pixels, it is hard to input the whole satellite image into a semantic segmentation neural network. Therefore, these satellite images are cropped to patches with a size of 768×768 pixels, resulting in a collection of 28,378 images.

In our dataset, the river ice images are segmented into

three broad categories, i.e., open water, river ice, and land. Moreover, according to the characteristics and development of river ice, the river ice is further classified into six fine-grained categories. Figure 2 shows some typical samples. The detailed description of each category is as follows.

1) Open water: It refers to areas in the image that are not frozen, representing the water surface or flowing water bodies. Open water exhibits a smoother texture compared to ice.

2) Shore ice: It is characterized by a smooth uniform surface, forms along river banks and channel bars. It is also

named as boarder ice or border ice in some literature. As shown in the first two rows of Figure 2, shore ices are marked by yellow. The shore ice along bank is usually in the shape of a strip, while the shore ice along channel bar is usually in the shape of a collar.

3) Drift ice: It refers to scattered ice floes that form during the floating ice period as a result of temperature drops. These ices float on the water surface and drift with the current. Drift ice may undergo different status, such as frazil slush, frazil ice pan and crusty pan along with temperature change. With drifting on the water surface, they will join/collide with other ice, small ice pans may join together to form big ice pan/strip. As shown in the first two rows of Figure 2, this kind of ice (marked by pink) has diverse shapes and sizes. The size ranges from several pixels to thousands of pixels.

4) Thermal growth ice: Typically, this type of ice forms in environments where water bodies are still or where the current is exceedingly slow. As the water temperature drops below the freezing point, the surface begins to freeze, resulting in the formation of a flat and continuous ice layer.

5) Consolidated ice blocks: This kind of ice forms in environments with rapid water flow, particularly in the fast currents or narrow sections of rivers. For example, drift ice blocks floe with the current. As the temperature decreases, with the amount of drift ice increases, they accumulate together so that the drift ice density continuously increases. When they reach the narrow river channel, they are stuck in the river channel and no longer drift, meanwhile if water cools quickly, they will freeze up and never drift to form a continuous large sheet of ice, i.e. consolidated ice block. The morphology of this ice is more chaotic, and it sometimes appears as though the ice blocks have been randomly stacked together.

6) Mixture of snow with ice or others: It refers to the mixture forms when snow and ice layers or others in the river channel combine after a snowfall. It exhibits higher brightness and reflectivity.

7) Other ice cover: It refers to ice that forms on land, typically near wetlands or adjacent areas to the river, but not directly connected to river channel. This type of ice may be formed by the freezing of water overflowing from the rising water level in the river channel.

8) Land: It refers to the other ground areas outside the river channel and not covered by ice.

C. Dataset Characteristics

To our best knowledge, the presented NWPU_YRCC_GFICE dataset is the first satellite image dataset for fine-grained river ice semantic segmentation. All images are captured by GF-2 satellite in Ningxia-Inner Mongolia section of Yellow River, in which river ice phenomenon is very typical. In total, it contains 28,378 images with a size of 768×768 pixels. Compared with the existing river ice datasets with semantic labels, it is currently the largest one. To fine explore the river ice start, development and evolution from satellite view, we carefully classify the river ice into six fine-grained subclasses at pixel-level. Only the dataset collected from the Dauphin River, presented by Ansari et al. [51], can be on a par with our

dataset in terms of the diversity of river ice types. Meanwhile, our dataset spans river ice conditions of the Yellow River (2015–2024) during freeze–thaw seasons, with image acquisitions concentrated between November and March, and the comparable dataset only contains 115 labeled images collected from November to March of three consecutive years (2017 to 2020). Therefore, our dataset has significant advantages in terms of data diversity. In conclusion, NWPU_YRCC_GFICE dataset enriches the river ice dataset, enabling fine-grained large-scale monitoring of river ice from satellite view.

IV. METHOD

To establish a comprehensive benchmark for fine-grained river ice segmentation, we propose two baseline models: ICEYOLO and ICEFormer. ICEYOLO is a CNN-based model built upon the YOLOv8 architecture, designed for real-time instance segmentation of river ice. In contrast, ICEFormer is a Transformer-based semantic segmentation model, featuring a hierarchical encoder-decoder structure to capture rich spatial and spectral information. In the following sections, we detail the architectural designs and key components of both models.

A. ICEYOLO

Since river ice monitoring tasks often require both high segmentation accuracy and real-time inference speed, ICEYOLO leverages the YOLOv8 framework and two improvements to balance precision and efficiency.

Although YOLOv8 performs well on many panoramic instance segmentation tasks with competitive inference speed advantage, the direct use is not suitable for river ice semantic segmentation. This is mainly attributed to two aspects: 1) The irregular shapes and fragmented boundaries commonly observed in river ice formations make standard instance label conversion prone to geometric distortion. As shown in the third column of Figure 4, the instance boundary generated by original YOLOv8 marked by red rectangle is incomplete or inaccurate, especially for instances on the image edge and small instances; 2) River ice often exhibits subtle textural variations across different ice types, such as shore ice, drift ice, and consolidated ice blocks.

To address the challenges posed by river ice segmentation, two targeted improvements are introduced: 1) Adaptive Key-point Redistribution Algorithm is proposed, ensuring more uniform and accurate boundary point distribution during instance generation. 2) Channel Attention Mechanism (CAM) [54] is incorporate into the neck part of YOLOv8 architecture, allowing the network to dynamically prioritize informative feature channels and improve segmentation accuracy under complex environmental conditions. The details of the entire model architecture and two improvements are introduced one by one as follows.

Model Architecture. Similar with YOLOv8, the ICEYOLO model follows a four-stage architecture: preprocessing, backbone, neck, and segmentation head, as shown in Figure 3(a). In the preprocessing stage, semantic segmentation masks are transformed into YOLO-compatible instance segmentation

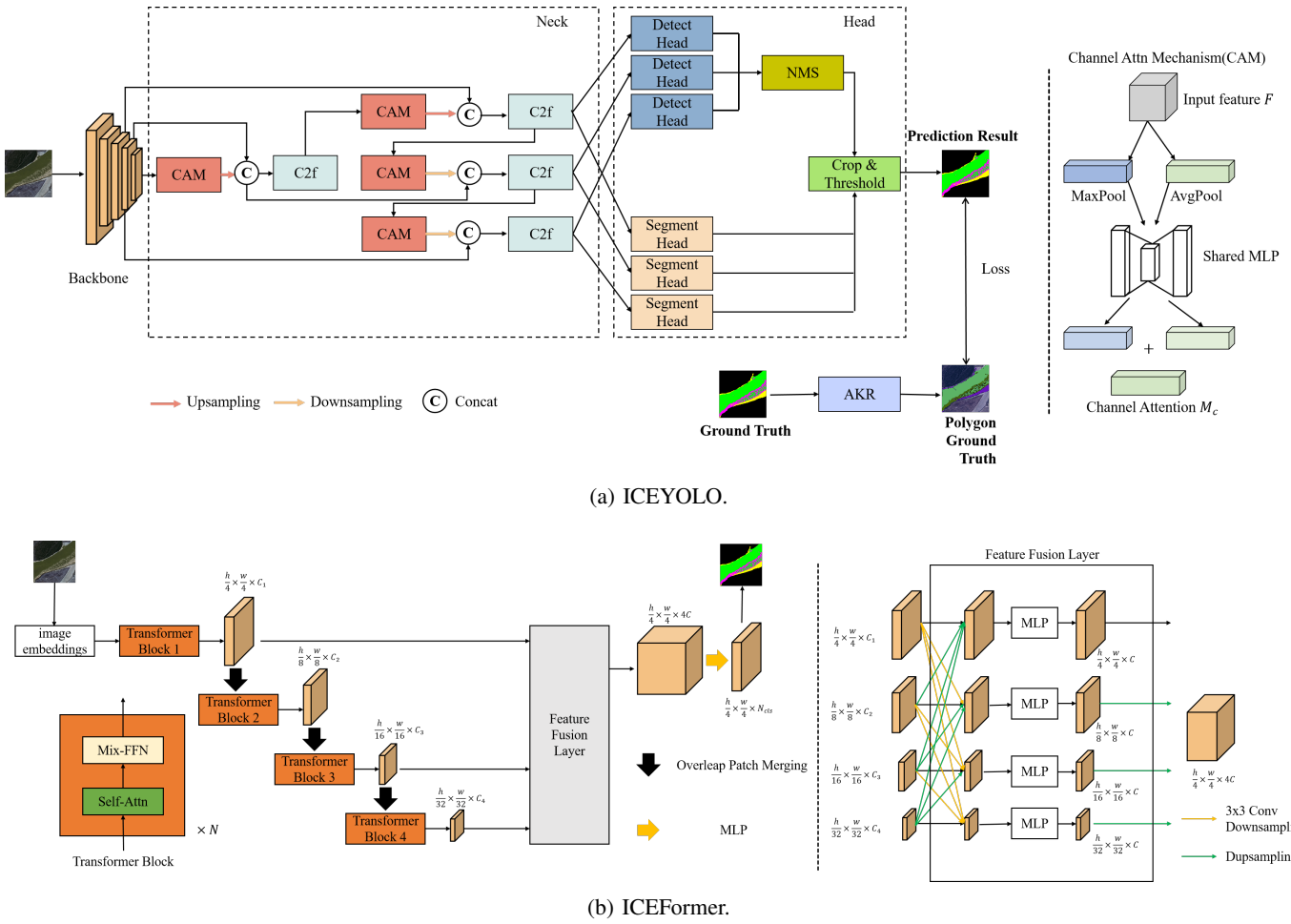


Fig. 3: The architectures of the proposed river ice segmentation models: (a) ICEYOLO, an enhanced YOLO-based river ice segmentation framework; (b) ICEFormer, a Transformer-based river ice segmentation network.

labels. This involves the use of the Teh-Chin chain approximation algorithm to extract boundary points, followed by the Adaptive Keypoint Redistribution algorithm (AKR) to optimize keypoint distribution, preserving boundary integrity and minimizing information loss. The backbone adopts the YOLOv8 feature pyramid design, which incorporates C2f modules and depthwise separable convolutions to efficiently extract multiscale hierarchical features from the input images. Feature maps are then aggregated in the neck module, where the Channel Attention Mechanism (CAM) is inserted before each C2f block in the multiscale feature aggregation path of the neck to selectively enhance critical feature representations. Finally, the segmentation head predicts instance masks, bounding boxes, and confidence scores for detected river ice objects.

The Adaptive Keypoint Redistribution Algorithm. During the transformation from semantic masks to instance labels, the uneven distribution of boundary points can severely distort the geometry of complex river ice structures, thereby impairing segmentation performance. This unevenness often arises when long or intricately shaped river ice objects produce mask contours where some segments are sparsely sampled while others are overly dense, leading to jagged or biased polygonal approximations. The original transformation algorithm in

YOLOv8 faces this issue when coping with river ice, as new keypoints are added without considering the overall shape and distribution of the boundaries.

To overcome this, the Adaptive Keypoint Redistribution Algorithm is designed to iteratively balance the distribution of keypoints along object boundaries. The algorithm first decomposes the boundary into segments and inserts them into a max-heap sorted by length. The longest segment is selected at each iteration for keypoint insertion, progressively achieving a uniform point distribution. This procedure continues until the target number of keypoints is reached. Finally, keypoints are collected and sorted to reconstruct a smooth and consistent boundary contour. Multi-threaded processing is adopted to accelerate computation. The procedural details are outlined in Algorithm 1.

Channel Attention Mechanism. The intricate textures and subtle inter-class variations characteristic of river ice present challenges for traditional convolutional feature extraction. To address this, a Channel Attention Mechanism (CAM) [54] is embedded into the neck module of ICEYOLO. Specifically, CAM modules are inserted before each C2f block in the multiscale feature aggregation path of the neck, enhancing channel-wise feature representations at three different scales. By

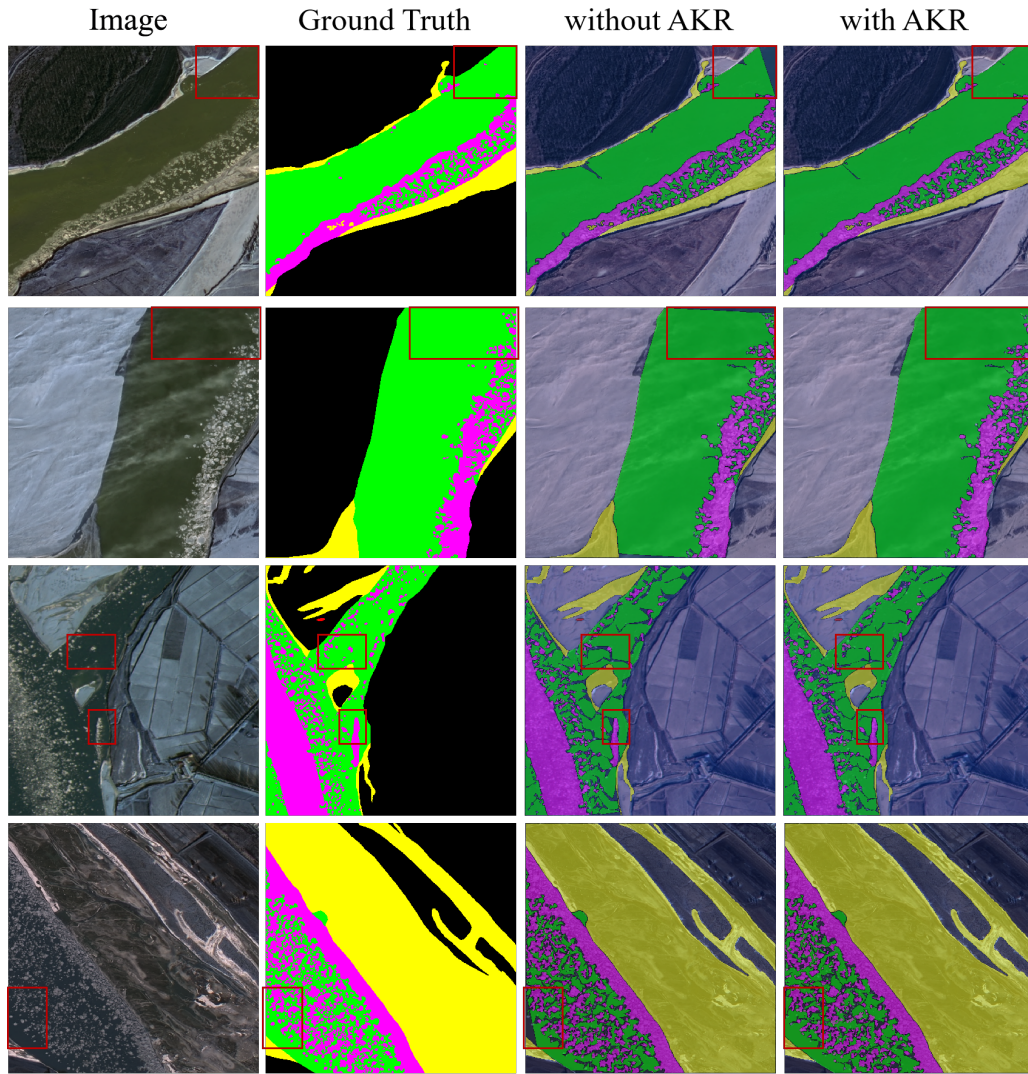


Fig. 4: Adaptive Keypoint Redistribution Algorithm. The original algorithm for adding keypoints in YOLOv8 often results in significant loss of important regions, as new keypoints are added without considering the overall shape and distribution of the boundaries. This can lead to poor representation of the river ice contours. In contrast, the proposed Adaptive Keypoint Redistribution Algorithm ensures a more balanced and accurate distribution of keypoints along the boundary, preserving the integrity of the original shape and minimizing the loss of crucial details.

dynamically adjusting feature weights, CAM assigns higher attention scores to informative channels while suppressing redundant responses. This improves the model's sensitivity to ice-related patterns such as thin drift ice and fragmented shore ice, ultimately boosting segmentation accuracy in complex river scenes. The structure of CAM is depicted by the right part of Figure 3(a).

B. ICEFormer

ICEFormer is a Segformer-based semantic segmentation benchmark designed for high-precision classification of river ice. Compared to CNN-based models, Transformer architecture offers stronger global context modeling, making them particularly suitable for capturing the complex spatial distributions of different river ice types over large areas. Considering the multiscale characteristics of river ice, i.e. the difference

in size of diverse river ice instances is huge, we improve the original Segformer by a well designed multiscale feature fusion module. The details of the ICEFormer architecture and feature fusion module are introduced as follows.

Model Architecture. ICEFormer adopts a hierarchical encoder-decoder architecture, the same as Segformer. The encoder directly processes the input remote sensing image through multiple Transformer stages, progressively downsampling the feature maps to 1/4, 1/8, and 1/16 resolution of the original input image. Each Transformer block integrates a multi-head self-attention mechanism to capture long-range dependencies and a Mix Feed-Forward Network (Mix FFN) to enhance local feature extraction via depthwise separable convolutions. This design enables ICEFormer to simultaneously model fine-grained ice features and large-scale spatial structures, addressing the characteristics of river ice scenes.

Algorithm 1 Polygon Boundary Keypoint Alignment

Require: segments (original boundary points of the polygon), n (target number of keypoints)

Ensure: points (aligned boundary keypoints)

```

1: Initialize a max-heap heap
2: total_point_count = len(segments) - 1
3: # Initialize the heap
4: for each pair of consecutive points in segments do
5:     length = distance between (point1.x, point1.y) and (point2.x, point2.y)
6:     line = create segment object:
7:     line.index = current segment index
8:     line.point1 = (point1.x, point1.y)
9:     line.point2 = (point2.x, point2.y)
10:    line.point_count = 1
11:    line.length = length
12:    push line into heap
13: end for
14: # Add points
15: while total_point_count <  $n$  do
16:    line = extract max-length segment from heap
17:    line.point_count += 1
18:    update line and push back into heap
19:    total_point_count += 1
20: end while
21: # Sort segments by original index
22: sorted_lines = sort segments in heap by original index
23: Initialize points array
24: # Collect points
25: for each line in sorted_lines do
26:    add line.point1 to points
27:    for  $i$  from 0 to line.point_count - 1 do
28:         $time = i / (line.point\_count - 1)$ 
29:         $new\_x = line.point1.x + time \times (line.point2.x - line.point1.x)$ 
30:         $new\_y = line.point1.y + time \times (line.point2.y - line.point1.y)$ 
31:        add (new_x, new_y) to points
32:    end for
33: end for
34: # Handle the last remaining point
35: add the last point of the original segments to points
36: return points

```

The decoder aggregates multiscale features by the designed feature fusion module and progressively reconstruct high-resolution segmentation maps. Finally, a MLP (Multilayer Perceptron) layer is utilized to refine the fused features, generating the output segmentation map with improved spatial consistency and class discrimination.

Feature Fusion Layer. To fully exploit multiscale semantic information, a dedicated feature fusion layer is introduced before final prediction. As illustrated in the right part of Figure 3(b), there are totally four groups of feature map with different scales, produced by our encoder. For each scale, features from other scales are through a resolution-aligned process: those from coarser scales are downsampled using 3×3 convolutions, and those from finer scales are upsampled; then, these resized feature maps are aggregated with themselves by a concatenation operation. To minimize interpolation artifacts

and preserve fine boundary details, data-dependent upsampling (DUpsampling) [55] is employed for upsampling during resolution restoration. Subsequently, the channel dimensions across all aggregated feature of different scales are unified via MLPs, followed by concatenation along the channel axis to produce the fused features.

V. EXPERIMENTS

A. Dataset

In addition to NWPU_YRCC_GFICE, we conduct experiments on three publicly available river ice segmentation datasets, i.e., NWPU_YRCC_EX, NWPU_YRCC2 and Alberta River Ice Segmentation Dataset, to assess the generalization capability of our models. For fair comparison, ICEYOLO and ICEFormer are trained and evaluated individually on each dataset. NWPU_YRCC_EX contains 887 pixel-wise annotated

TABLE II: Benchmark on NWPU_YRCC_GFICE dataset. CNN-based models and Transformer-based models are separated for comparison.

Method	IoU(%)								mIoU(%)	FPS
	Thermal Growth Ice	Consolidated Ice Blocks	Shore Ice	Drift Ice	Mixture of snow with ice or others	Water	Other Ice Cover	Land		
CNN-based Models										
FCN [4]	55.00	57.42	58.83	57.42	70.11	81.96	93.28	33.26	65.60	25.92
UNet [9]	64.12	65.56	49.18	61.09	58.15	78.29	65.30	89.96	66.45	22.39
DeepLabV3 [7]	75.03	67.29	62.17	56.75	65.61	82.34	76.70	94.95	72.61	20.52
PSPNet [56]	68.63	55.50	46.88	55.22	25.41	78.77	70.01	91.57	61.50	27.02
SegNet [26]	39.66	31.19	1.00	12.89	11.27	59.66	43.08	85.16	35.49	20.98
ERFNet [57]	29.32	44.12	69.59	44.12	67.50	79.75	91.93	14.20	55.31	92.58
RefineNet [12]	43.18	41.94	16.37	28.10	18.93	68.24	55.20	88.23	58.75	18.10
UNet++ [10]	57.02	49.94	2.98	19.25	15.53	72.88	60.21	88.53	45.79	19.81
DeepLabV3+ [8]	72.73	67.20	54.35	58.71	44.33	83.84	76.89	94.71	69.09	24.22
PSANet [58]	62.66	46.62	46.73	54.43	15.55	80.11	68.56	89.77	58.05	19.95
DenseASPP [59]	74.80	64.80	56.21	54.80	40.01	83.53	76.74	94.24	68.14	55.00
BiseNet [15]	56.89	60.08	72.31	60.08	78.81	86.30	94.67	55.43	72.95	148.62
ENCNet [60]	51.73	64.48	57.62	51.89	72.91	82.24	93.99	19.40	66.38	27.31
HRNet [61]	53.53	56.11	77.35	56.11	72.53	86.31	93.60	30.42	66.63	22.27
CCNet [62]	76.68	64.55	64.97	62.94	40.07	85.59	78.46	93.23	70.81	24.09
SemFPN [63]	67.41	39.99	45.36	59.51	64.12	78.52	72.73	94.59	65.28	70.39
APCNet [64]	81.06	65.13	69.26	73.21	50.42	86.12	81.75	94.75	75.21	23.11
ANN [65]	85.54	66.33	79.11	83.57	74.63	87.67	84.68	95.99	82.19	25.80
DANet [66]	84.30	65.70	76.12	79.64	55.99	88.23	83.78	95.39	78.64	21.70
OCRNet [67]	77.78	68.69	66.47	60.73	60.86	84.74	78.77	94.92	74.12	30.66
DDRNet [68]	59.33	42.79	40.01	46.51	37.36	65.59	66.46	91.19	56.15	109.61
KNet [69]	75.11	61.29	65.36	64.82	44.48	86.64	79.46	94.52	71.46	24.89
BiseNetV2 [70]	51.86	32.74	19.30	16.81	16.75	36.14	49.93	82.90	38.30	129.39
STDC [71]	73.40	63.27	58.74	49.45	49.27	77.57	68.49	92.91	66.64	122.47
SegNext [72]	84.43	64.89	75.46	83.50	80.49	87.28	83.21	96.03	81.91	63.79
PIDNet [16]	66.70	56.70	53.59	58.25	53.31	70.11	60.20	84.90	62.97	88.08
ICEYOLO (Ours)	83.66	82.45	76.85	69.19	78.52	86.70	85.88	96.65	82.48	39.21
Transformer-based Models										
ViT [73]	79.95	56.78	72.93	76.51	77.17	85.95	79.65	95.54	78.06	8.94
SETR [14]	53.64	41.29	30.64	28.89	49.89	62.71	61.76	91.49	52.54	5.87
DPT [74]	74.41	71.71	51.31	50.05	64.99	83.35	77.28	95.32	71.05	12.23
Segmenter [75]	61.49	49.63	27.20	19.02	52.47	65.85	65.98	93.38	54.38	14.89
SegFormer [13]	84.22	84.64	78.19	71.96	80.75	89.02	82.71	96.10	83.45	16.71
Swin [36]	77.19	70.83	80.28	70.83	83.89	88.74	95.88	79.25	82.31	20.20
TWINS [76]	71.89	66.52	77.78	66.52	80.97	88.40	95.81	75.17	79.68	38.24
ICEFormer (Ours)	84.23	85.69	78.21	72.34	82.04	89.30	82.72	96.10	83.83	15.57

UAV images, splitting into 524 training images, 180 validation images, and 183 test images. The NWPU_YRCC2 dataset comprises 1,525 annotated images and follows a 3:1:1 split, resulting in 915 images for training, 305 for validation, and 305 for testing. The Alberta River Ice Segmentation Dataset is generated by cropping and augmenting 50 manually annotated UAV and camera-captured images, resulting in 554 training images and 138 testing images.

B. Evaluation Criteria

In our experiments, we adopt two key metrics: mean Intersection over Union (mIoU) to assess segmentation accuracy, and Frames Per Second (FPS) to reflect inference efficiency. These two metrics together provide a comprehensive evaluation

of each model's balance between precision and speed, which is particularly important for real-time monitoring and large-scale satellite image analysis.

C. Experimental Settings

To ensure a fair and rigorous comparison, all models in our benchmark are trained individually with hyperparameter tuning tailored to each architecture. Segmentation heads are adapted to match the number of target classes in each dataset. All input images are resized to a resolution of 768×768 pixels. Segmentation heads are modified as necessary to accommodate the class numbers specific to each dataset. All hyperparameters have been systematically optimized through

validation experiments to achieve optimal performance for each model architecture.

ICEYOLO Training Settings. ICEYOLO is optimized using the Adam optimizer with an initial learning rate of 0.01. The learning rate follows a linear decay schedule over 500 training epochs. A batch size of 40 is used during training. Random image flipping is applied for data augmentation. The Adaptive Keypoint Redistribution Algorithm is applied to each polygon instance, with the target number of keypoints set to 1,000, which was experimentally determined to be optimal.

ICEFormer Training Settings. ICEFormer is trained using the AdamW optimizer, with an initial learning rate of $6e-5$ and a weight decay of 0.01. A polynomial learning rate decay policy is adopted. Training is conducted for 85 epochs, with a batch size of 12. Data augmentation includes random scaling within the range [0.5, 2.0] and random horizontal flipping.

Hardware Environment. All experiments are implemented using PyTorch and conducted on four NVIDIA GeForce RTX 4090 GPUs. Inference speed (Frames Per Second, FPS) is measured separately on a single NVIDIA RTX 3090 GPU for consistency across all models.

Loss Functions. ICEYOLO adopts the original segmentation loss function of YOLO, and ICEFormer inherits the pixel-wise cross-entropy loss from SegFormer.

D. Benchmark on NWPU_YRCC_GFICE Dataset

Table II presents the benchmarking results of 33 existing models and the two proposed models, ICEYOLO and ICEFormer, on the NWPU_YRCC_GFICE dataset. Models are categorized into CNN-based and Transformer-based groups for a more structured comparison.

Among CNN-based methods, DeepLabV3+, OCRNet, and SegNext achieve relatively high mIoU scores, demonstrating strong segmentation capabilities. ICEYOLO, based on the YOLOv8 architecture and improved with adaptive keypoint redistribution and multiscale attention mechanisms, achieves a superior mIoU of 82.48% while maintaining a high inference speed of 39.21 FPS. This balance of accuracy and efficiency highlights the advantages of ICEYOLO for practical, river ice monitoring applications. Conversely, SegNet exhibits the poorest performance among CNN-based methods with only 35.49% mIoU. SegNet's poor performance is attributed to its reliance on max-pooling indices transmission and Max-Unpool2d operations, leading to significant spatial detail loss and limited boundary reconstruction capability for the complex spectral and geometric patterns in satellite-based river ice imagery.

For Transformer-based methods, SegFormer, Swin Transformer, and ViT exhibit competitive segmentation performance, leveraging their strong global context modeling capabilities. Notably, the proposed ICEFormer achieves an mIoU of 83.83%, outperforming SegFormer and other Transformer baselines. ICEFormer also maintains reasonable inference efficiency with 15.57 FPS, validating the effectiveness of the multiscale feature fusion and hierarchical enhancements tailored for fine-grained river ice segmentation. In contrast, SETR exhibits the poorest performance among Transformer-based methods with only 52.54% mIoU. SETR's suboptimal

performance stems from the absence of multi-scale feature fusion mechanisms and large patch usage, which results in fine-grained detail loss and blurred boundaries.

In summary, ICEFormer achieves superior segmentation accuracy, outperforming all baseline methods, whereas ICEYOLO maintains a favorable balance between segmentation accuracy and real-time inference efficiency. The results demonstrate that the architectural adaptations and preprocessing strategies introduced in this study effectively address the challenges posed by temporal variability and morphological complexity in satellite-based river ice segmentation.

E. Comparison Experiments

To assess model generalization, we further evaluate ICEYOLO and ICEFormer on three public datasets: NWPU_YRCC_EX, NWPU_YRCC2, and Alberta. As shown in Table III, both models outperform all other 14 segmentation models across all datasets.

ICEYOLO achieves the top mIoU score on NWPU_YRCC2 (92.34%), while ICEFormer leads on NWPU_YRCC_EX with the mIoU score 92.98% and on Alberta River Ice Segmentation dataset with the mIoU score 96.82%. These results demonstrate that our models perform well on other river ice datasets.

F. Ablation Studies

To validate the effectiveness of individual components in our proposed models, we conduct comprehensive ablation studies on the NWPU_YRCC_GFICE dataset. These experiments systematically evaluate the contribution of each proposed module by progressively adding components to the baseline architectures. Table IV presents the quantitative results of this ablation study.

ICEYOLO Component Analysis. For ICEYOLO, we evaluate the individual and combined contributions of the Adaptive Keypoint Redistribution algorithm (AKR) and Channel Attention Mechanism (CAM). The baseline configuration uses the original YOLOv8 architecture without any modifications. Figure 5 (left) visualizes the comparative results between YOLOv8 and ICEYOLO.

The baseline achieves an mIoU of 80.4%. The incorporation of AKR alone yields a performance improvement of 0.68%, reaching 81.08% mIoU. This improvement demonstrates that the proposed keypoint redistribution strategy effectively preserves boundary integrity during the semantic-to-instance label conversion process, particularly benefiting the segmentation of irregularly shaped river ice formations.

The incorporation of CAM alone elevates the mIoU to 82.1%, reflecting a notable improvement of 1.7% over the baseline. This significant enhancement suggests that the CAM effectively captures subtle textural differences among various ice types, allowing the model to classify pixels more accurately based on their contextual information. Furthermore, AKR module enhances segmentation precision for complex and irregularly shaped ice formations. When both AKR and CAM are integrated into the full ICEYOLO framework, the model attains its highest performance, reaching 82.48% mIoU. The

TABLE III: Performance comparison on three public river ice segmentation datasets: NWPU_YRCC_EX, NWPU_YRCC2, and the Alberta River Ice Segmentation dataset. Our proposed models (ICEYOLO and ICEFormer) are highlighted in bold. The best-performing results for each dataset are marked in red, while the second-best results are marked in blue.

Method	mIoU (%)		
	NWPU_YRCC_EX [29]	NWPU_YRCC2 [29]	Alberta River Ice Segmentation dataset [24]
ENet [77]	89.94	81.37	80.75
CGNet [78]	90.93	80.49	81.62
DABNet [79]	90.05	80.12	82.13
FPENet [80]	83.61	76.89	79.89
FSSNet [81]	89.15	79.63	78.04
LEDNet [82]	91.27	81.81	81.56
ContextNet [83]	87.16	77.83	78.64
FastSCNN [84]	87.13	77.36	78.11
ERFNet [57]	91.57	81.58	81.94
LinkNet [85]	90.45	81.29	82.12
BiSeNet [15]	89.72	80.09	81.24
STDC-2 [86]	91.18	81.46	81.86
PP-LiteSeg-T2 [87]	89.54	79.99	80.39
PIDNet-S [16]	90.21	80.29	81.27
PIDNet-L [16]	91.25	81.22	82.10
ICEYOLO (Ours)	88.65	92.34	93.17
ICEFormer (Ours)	92.98	90.89	96.82

TABLE IV: Ablation Study Results on NWPU_YRCC_GFICE Dataset

Method	mIoU (%)
ICEYOLO Variants	
Baseline (YOLOv8)	80.40
+ AKR	81.08
+ CAM	82.10
+ AKR + CAM (ICEYOLO)	82.48
ICEFormer Variants	
Baseline (SegFormer)	83.45
+ FFL (ICEFormer)	83.83

AKR: Adaptive Keypoint Redistribution;
CAM: Channel Attention Mechanism;
FFL: Feature Fusion Layer

superiority of this combined approach is further illustrated in the qualitative comparisons shown in Figure 5 (left), highlighting the improved delineation and class-specific accuracy achieved by the proposed modules relative to the baseline.

ICEFormer Feature Fusion Analysis. For ICEFormer, we evaluate the contribution of the proposed Feature Fusion Layer (FFL) by comparing the enhanced model against the standard SegFormer baseline. The baseline SegFormer architecture achieves 83.45% mIoU on our dataset, establishing a strong foundation for transformer-based segmentation. Figure 5 (right) visualizes the comparative results between Segformer and ICEFormer.

The incorporation of the multiscale Feature Fusion Layer increases the performance to 83.83% mIoU, representing an improvement of 0.38%. While this enhancement appears modest in absolute terms, it is statistically significant and demonstrates the value of explicitly modeling multiscale feature relationships for fine-grained river ice classification. By fusing features across different dimensions, the Feature Fusion Layer

substantially improves the model's capacity to segment classes with higher accuracy, effectively capturing ice formations of various scales from small drift ice fragments to large consolidated blocks. This improvement is further corroborated by the qualitative results presented in Figure 5 (right), which illustrate the enhanced discriminative capability enabled by this fusion strategy.

VI. CONCLUSION

In this study, we address the critical gap in fine-grained river ice segmentation using satellite image by introducing the NWPU_YRCC_GFICE dataset. This dataset constructed from multi-spectral GF-2 satellite images, provides pixel-level annotations across eight detailed categories and spans multiple freeze-thaw cycles of the Yellow River. It is the first publicly available satellite-based dataset specifically designed for fine-grained river ice classification.

To benchmark segmentation performance, we develop two tailored models: ICEYOLO and ICEFormer. ICEYOLO, based on the YOLOv8 architecture with targeted enhancements including Adaptive Keypoint Redistribution and Channel Attention Mechanism, achieves an excellent balance between segmentation accuracy and inference speed. ICEFormer, built upon the SegFormer framework with multiscale feature fusion, achieves the highest segmentation accuracy among all evaluated methods. Extensive experiments demonstrate that both models outperform numerous existing methods on the NWPU_YRCC_GFICE dataset, while maintaining strong generalization capability across three public river ice datasets.

The NWPU_YRCC_GFICE dataset and the proposed models provide benchmarks for advancing satellite-based river ice monitoring. Future work will explore more efficient Transformer architectures and self-supervised learning strategies

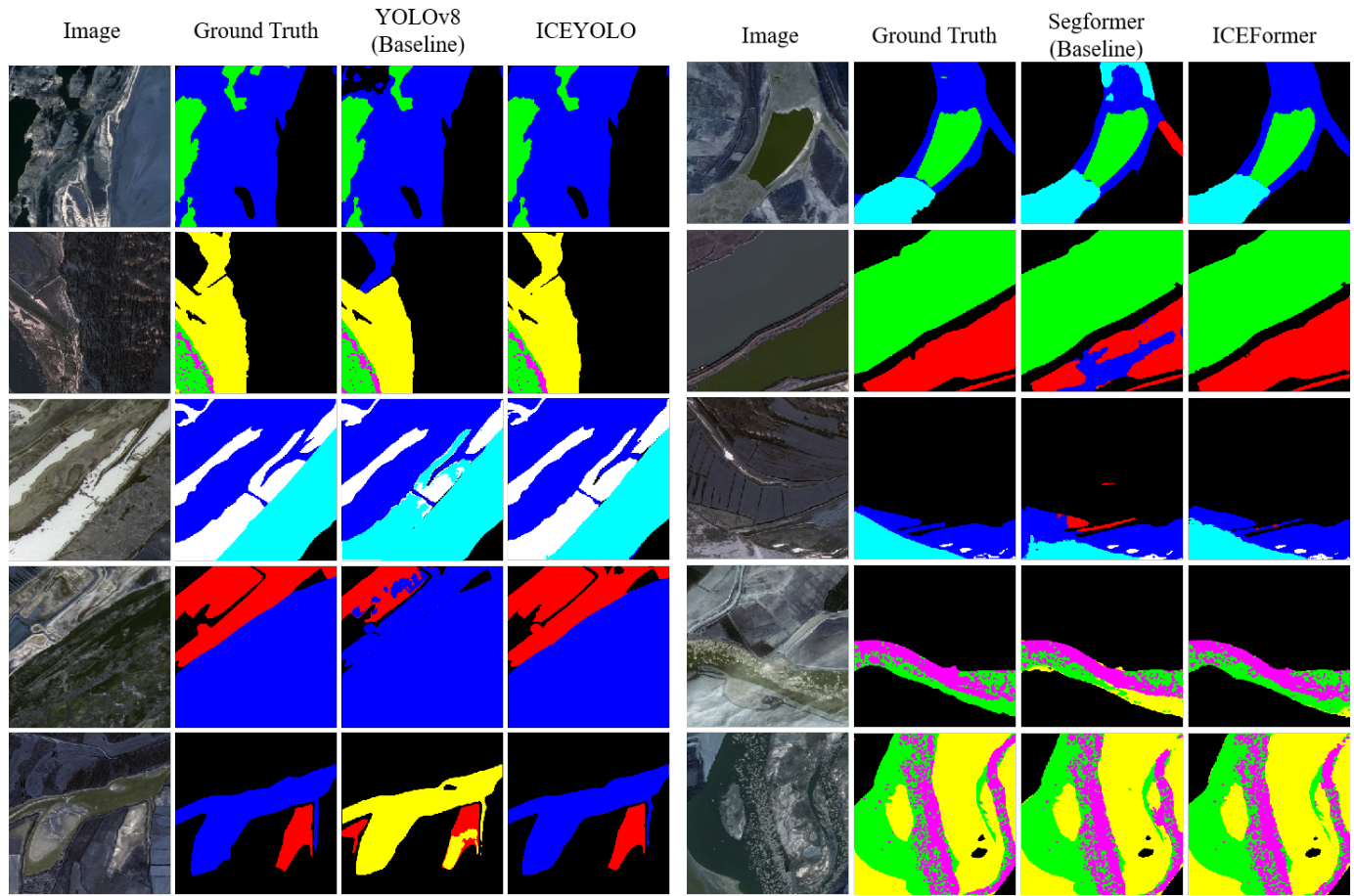


Fig. 5: Visual comparison of YOLOv8 vs ICEYOLO (left) and Segformer vs ICEFormer (right) on NWPU_YRCC_GFICE dataset.

to further enhance segmentation performance under limited labeled data conditions.

REFERENCES

- [1] I. Zakharov, T. Puestow, A. A. Khan, R. Briggs, and P. Barrette, "Review of river ice observation and data analysis technologies," *Hydrology*, vol. 11, no. 126, 2024.
- [2] B. Altena and A. Kääb, "Quantifying river ice movement through a combination of european satellite monitoring services," *International Journal of Applied Earth Observation and Geoinformation*, vol. 98, p. 102315, 2021.
- [3] A. Sharifi, "Development of a method for flood detection based on sentinel-1 images and classifier algorithms," *Water and Environment Journal*, vol. 35, no. 3, pp. 924–929, 2021.
- [4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [5] L.-C. Chen, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.
- [6] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
- [7] L.-C. Chen, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.
- [8] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

- [10] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*. Springer, 2018, pp. 3–11.
- [11] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang *et al.*, "Deep high-resolution representation learning for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3349–3364, 2020.
- [12] G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1925–1934.
- [13] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in neural information processing systems*, vol. 34, pp. 12077–12090, 2021.
- [14] S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, T. Xiang, P. H. Torr *et al.*, "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 6881–6890.
- [15] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 325–341.
- [16] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "Pidnet: A real-time semantic segmentation network inspired by pid controllers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 19529–19539.
- [17] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolov8," 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [18] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015–4026.
- [19] X. Zhao, W. Ding, Y. An, Y. Du, T. Yu, M. Li, M. Tang, and J. Wang, "Fast segment anything," *arXiv preprint arXiv:2306.12156*, 2023.
- [20] C. Zhang, D. Han, Y. Qiao, J. U. Kim, S.-H. Bae, S. Lee, and C. S. Hong, "Faster segment anything: Towards lightweight sam for mobile applications," *arXiv preprint arXiv:2306.14289*, 2023.
- [21] Y. Xiong, B. Varadarajan, L. Wu, X. Xiang, F. Xiao, C. Zhu, X. Dai, D. Wang, F. Sun, F. Iandola *et al.*, "Efficientsam: Leveraged masked image pretraining for efficient segment anything," *arXiv preprint arXiv:2312.00863*, 2023.
- [22] Z. Yang, Y. Zhu, X. Zeng, J. Zong, X. Liu, R. Tao, X. Cong, and Y. Yu, "An easy zero-shot learning combination: Texture sensitive semantic segmentation ichrnet and advanced style transfer learning strategy," *arXiv preprint arXiv:2310.00310*, 2023.
- [23] X. Zhang, J. Jin, Z. Lan, C. Li, M. Fan, Y. Wang, X. Yu, and Y. Zhang, "Icenet: A semantic segmentation deep network for river ice by fusing positional and channel-wise attentive features," *Remote Sensing*, vol. 12, no. 2, p. 221, 2020.
- [24] A. Singh, H. Kalke, M. Loewen, and N. Ray, "River ice segmentation with deep learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 7570–7579, 2020.
- [25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
- [26] Z. Wu, C. Shen, and A. v. d. Hengel, "Real-time semantic image segmentation via spatial sparsity," *arXiv preprint arXiv:1712.00213*, 2017.
- [27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [28] X. Zhang, Y. Zhou, J. Jin, Y. Wang, M. Fan, N. Wang, and Y. Zhang, "Icenv2: A fine-grained river ice semantic segmentation network based on uav images," *Remote Sensing*, vol. 13, no. 4, p. 633, 2021.
- [29] X. Zhang, Z. Zhao, L. Ran, Y. Xing, W. Wang, Z. Lan, H. Yin, H. He, Q. Liu, B. Zhang *et al.*, "Fasticenet: A real-time and accurate semantic segmentation model for aerial remote sensing river ice image," *Signal Processing*, vol. 212, p. 109150, 2023.
- [30] C. Pei, Y. She, and M. Loewen, "Deep learning based river surface ice quantification using a distant and oblique-viewed public camera," *Cold Regions Science and Technology*, vol. 206, p. 103736, 2023.
- [31] Z. Yang, J. Zong, Y. Zhu, X. Liu, R. Tao, and Y. Yu, "River ice regime recognition based on deep learning: Ice concentration, area, and velocity," *Water*, vol. 16, no. 1, p. 58, 2024.
- [32] M. Temimi, M. Abdelkader, A. Tounsi, N. Chaouch, S. Carter, B. Sjoberg, A. Macneil, and N. Bingham-Maas, "An automated system to monitor river ice conditions using visible infrared imaging radiometer suite imagery," *Remote Sensing*, vol. 15, no. 20, p. 4896, 2023.
- [33] P. Muhammad, C. Duguay, and K.-K. Kang, "Monitoring ice break-up on the mackenzie river using modis data," *The Cryosphere*, vol. 10, no. 2, pp. 569–584, 2016.
- [34] G. Marthandavilakom Prakasam *et al.*, "Detecting ice jams on the rivers in northern finland using sentinel-1," Master's thesis, 2022.
- [35] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [36] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10012–10022.
- [37] Z. Xu, J. Geng, and W. Jiang, "Mmt: Mixed-mask transformer for remote sensing image semantic segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.
- [38] C. Peng, K. Zhang, Y. Ma, and J. Ma, "Cross fusion net: A fast semantic segmentation network for small-scale semantic information capturing in aerial scenes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2021.
- [39] Y. Qiao, B. Zhong, B. Du, H. Cai, J. Jiang, Q. Liu, A. Yang, J. Wu, and X. Wang, "Sam enhanced semantic segmentation for remote sensing imagery without additional training," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–16, 2025.
- [40] C. M. Surdu, C. R. Duguay, H. K. Pour, and L. C. Brown, "Ice freeze-up and break-up detection of shallow lakes in northern alaska with spaceborne sar," *Remote Sensing*, vol. 7, no. 5, pp. 6133–6159, 2015.
- [41] A. Beaton, R. Whaley, K. Corston, and F. Kenny, "Identifying historic river ice breakup timing using modis and google earth engine in support of operational flood monitoring in northern ontario," *Remote sensing of environment*, vol. 224, pp. 352–364, 2019.
- [42] N. Chaouch, M. Temimi, P. Romanov, R. Cabrera, G. McKillop, and R. Khanbilvardi, "An automated algorithm for river ice monitoring over the susquehanna river using the modis data," *Hydrological Processes*, vol. 28, no. 1, pp. 62–73, 2014.
- [43] S. W. Cooley and T. M. Pavelsky, "Spatial and temporal patterns in arctic river ice breakup revealed by automated ice detection from modis imagery," *Remote sensing of environment*, vol. 175, pp. 310–322, 2016.
- [44] P. Bharathi and P. Subashini, "Texture based color segmentation for infrared river ice images using k-means clustering," in *2013 International Conference on Signal Processing, Image Processing & Pattern Recognition*. IEEE, 2013, pp. 298–302.
- [45] M. Tom, U. Kälin, M. Sütterlin, E. Baltsavias, and K. Schindler, "Lake ice detection in low-resolution optical satellite images," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 4, pp. 279–286, 2018.
- [46] X. Yang, T. M. Pavelsky, and G. H. Allen, "The past and future of global river ice," *Nature*, vol. 577, no. 7788, pp. 69–73, 2020.
- [47] D. Purnell, M. Daboor, P. Matte, D. Peters, F. Ancil, T. Ghobrial, and A. Pierre, "Observations of river ice breakup using gnss-ir, sar and machine learning," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2024.
- [48] S. Ansari, C. Rennie, S. Clark, and O. Seidou, "Icemarknet: River ice detection and characterization using deep learning algorithms applied to aerial photography," *Cold Regions Science and Technology*, vol. 189, p. 103324, 2021.
- [49] B. Fu, X. Sun, S. Ma, X. Ma, and Z. Liu, "Cseu-net: Convnext-se-u-net for river ice floe segmentation using unmanned aerial vehicle grayscale remote sensing images," *Journal of Applied Remote Sensing*, vol. 18, no. 4, pp. 046505–046505, 2024.
- [50] W. Zhao, Y. Xue, F. Han, X. Peng, Y. Zhao, J. Zhang, J. Yang, Q. Lin, and Y. Wu, "Research on semantic segmentation algorithm of high latitude urban river ice based on deep transfer learning," *International Journal of Remote Sensing*, pp. 1–22, 2024.
- [51] S. Ansari, C. Rennie, S. Clark, and O. Seidou, "River ice detection and classification using oblique shore-based photography," *Cold Regions Science and Technology*, vol. 228, p. 104303, 2024.

- [52] M. Ayyad, M. Temimi, M. Abdelkader, M. M. Henein, F. L. Engel, R. R. Löttspeich, and J. R. Eggleston, "Rice-net: Integrating ground-based cameras and machine learning for automated river ice detection," *Environmental Modelling & Software*, p. 106454, 2025.
- [53] W. Zhao, F. Han, Y. Xue, J. Zhang, Y. Zhao, X. Peng, and Q. Lin, "Improved ice channel segmentation and width calculation using drone-based optical sensors," *IEEE Sensors Journal*, 2023.
- [54] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
- [55] Z. Tian, T. He, C. Shen, and Y. Yan, "Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3126–3135.
- [56] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.
- [57] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo, "Erfnet: Efficient residual factorized convnet for real-time semantic segmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 263–272, 2017.
- [58] H. Zhao, Y. Zhang, S. Liu, J. Shi, C. C. Loy, D. Lin, and J. Jia, "Psanet: Point-wise spatial attention network for scene parsing," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 267–283.
- [59] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "Denseaspp for semantic segmentation in street scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3684–3692.
- [60] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, and A. Agrawal, "Context encoding for semantic segmentation," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018, pp. 7151–7160.
- [61] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5693–5703.
- [62] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnet: Criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 603–612.
- [63] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 6399–6408.
- [64] J. He, Z. Deng, L. Zhou, Y. Wang, and Y. Qiao, "Adaptive pyramid context network for semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 7519–7528.
- [65] Z. Zhu, M. Xu, S. Bai, T. Huang, and X. Bai, "Asymmetric non-local neural networks for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 593–602.
- [66] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3146–3154.
- [67] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*. Springer, 2020, pp. 173–190.
- [68] Y. Hong, H. Pan, W. Sun, and Y. Jia, "Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes," *arXiv preprint arXiv:2101.06085*, 2021.
- [69] W. Zhang, J. Pang, K. Chen, and C. C. Loy, "K-net: Towards unified image segmentation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 10 326–10 338, 2021.
- [70] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, and N. Sang, "Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation," *International journal of computer vision*, vol. 129, pp. 3051–3068, 2021.
- [71] M. Fan, S. Lai, J. Huang, X. Wei, Z. Chai, J. Luo, and X. Wei, "Rethinking bisenet for real-time semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 9716–9725.
- [72] M.-H. Guo, C.-Z. Lu, Q. Hou, Z. Liu, M.-M. Cheng, and S.-M. Hu, "Segnext: Rethinking convolutional attention design for semantic segmentation," *Advances in neural information processing systems*, vol. 35, pp. 1140–1156, 2022.
- [73] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [74] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 12 179–12 188.
- [75] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, "Segmenter: Transformer for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 7262–7272.
- [76] X. Chu, Z. Tian, Y. Wang, B. Zhang, H. Ren, X. Wei, H. Xia, and C. Shen, "Twins: Revisiting the design of spatial attention in vision transformers," *Advances in neural information processing systems*, vol. 34, pp. 9355–9366, 2021.
- [77] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," *arXiv preprint arXiv:1606.02147*, 2016.
- [78] T. Wu, S. Tang, R. Zhang, J. Cao, and Y. Zhang, "Cgnet: A light-weight context guided network for semantic segmentation," *IEEE Transactions on Image Processing*, vol. 30, pp. 1169–1179, 2020.
- [79] G. Li, I. Yun, J. Kim, and J. Kim, "Dabnet: Depth-wise asymmetric bottleneck for real-time semantic segmentation," *arXiv preprint arXiv:1907.11357*, 2019.
- [80] M. Liu and H. Yin, "Feature pyramid encoding network for real-time semantic segmentation," *arXiv preprint arXiv:1909.08599*, 2019.
- [81] X. Zhang, Z. Chen, Q. J. Wu, L. Cai, D. Lu, and X. Li, "Fast semantic segmentation for scene perception," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 1183–1192, 2018.
- [82] Y. Wang, Q. Zhou, J. Liu, J. Xiong, G. Gao, X. Wu, and L. J. Latecki, "Lednet: A lightweight encoder-decoder network for real-time semantic segmentation," in *2019 IEEE international conference on image processing (ICIP)*. IEEE, 2019, pp. 1860–1864.
- [83] W. Han, Z. Zhang, Y. Zhang, J. Yu, C.-C. Chiu, J. Qin, A. Gulati, R. Pang, and Y. Wu, "Contextnet: Improving convolutional neural networks for automatic speech recognition with global context," *arXiv preprint arXiv:2005.03191*, 2020.
- [84] R. P. Poudel, S. Liwicki, and R. Cipolla, "Fast-scnn: Fast semantic segmentation network," *arXiv preprint arXiv:1902.04502*, 2019.
- [85] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE visual communications and image processing (VCIP)*. IEEE, 2017, pp. 1–4.
- [86] M. Fan, S. Lai, J. Huang, X. Wei, Z. Chai, J. Luo, and X. Wei, "Rethinking bisenet for real-time semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 9716–9725.
- [87] J. Peng, Y. Liu, S. Tang, Y. Hao, L. Chu, G. Chen, Z. Wu, Z. Chen, Z. Yu, Y. Du *et al.*, "Pp-liteseg: A superior real-time semantic segmentation model," *arXiv preprint arXiv:2204.02681*, 2022.