Adapting Vision GNN for Patch-level Classification of Antarctic SAR Iceberg Imagery

Olivia Patterson, Thanh Nam Tran, Faith Kalendek, Joe Gung, Sree Nitya Kollu

Introduction

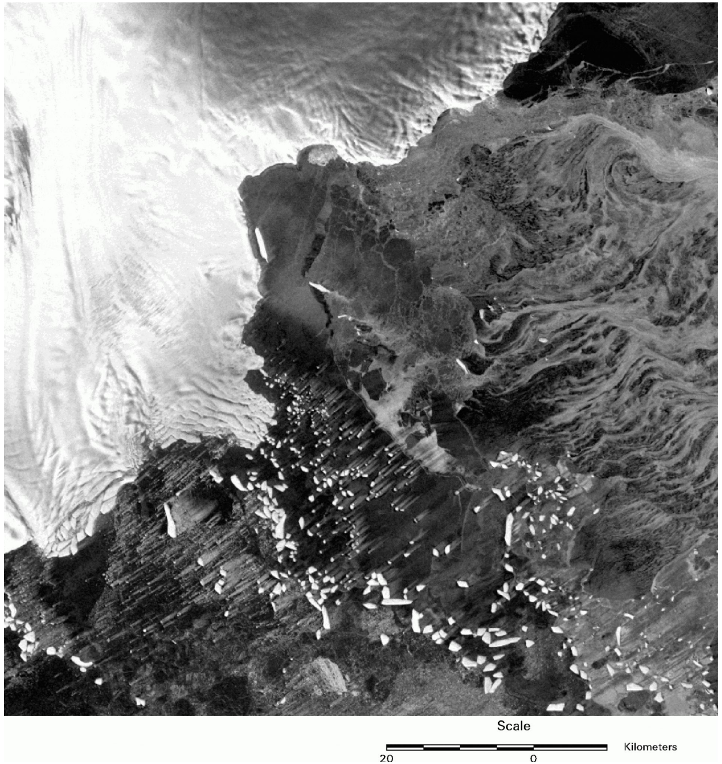
Antarctic Iceberg Imagery

- Understanding glacier and iceberg-related processes is important
  - Improve climate modeling & prediction
  - Inform policy decision
  - Icebergs calving and melting directly influence global sea-level rise
  - Polar systems have a critical role in regulating Earth's climate

Reference: https://www.thearcticinstitute.org/nasa-images-change-arctic/

Synthetic Aperture Radar

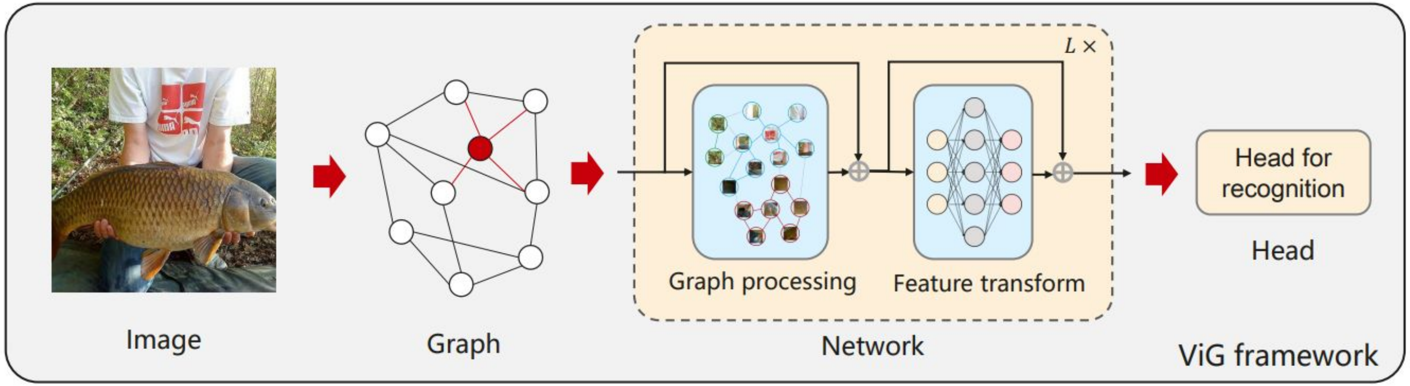
- Synthetic Aperture Radar (SAR)
  - Imaging radar system mounted on moving platforms (like satellites or aircraft)
  - Operates day or night and penetrates clouds, smoke, and light foliage
  - Efficient in variable atmospheric and lighting situations
  - A great tool for monitoring icebergs



Reference: https://www.earthdata.nasa.gov/news/sea-ice-motion-science-driving-u.s.-sar-research

Introduction

- Vision Graph Neural Network (ViG), a graph-based alternative for visual representation learning
  - Splits images into patches, treats patches as nodes, connects nearest neighbors, and exchanges information through graph-based modules

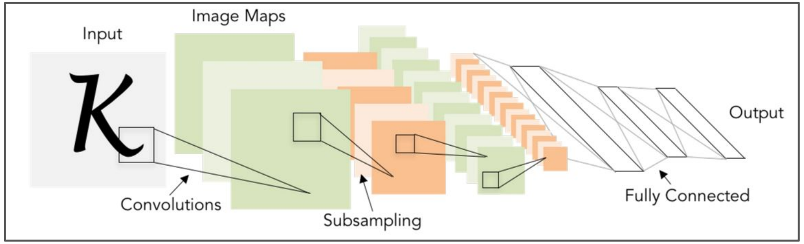


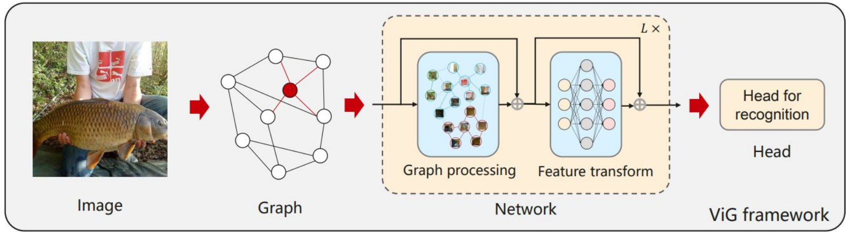
Reference: *Vision GNN: An Image is Worth Graph of Nodes*

Introduction

- Why ViG instead of CNN? Can we even use ViG for SAR iceberg imagery?
    - CNNs uses fixed grid-based kernels which bias the model toward local texture and regular spatial patterns.
    - However, patterns of SAR iceberg images (ice, ocean, etc.) can be irregular and fragmented.
    - Small ice fragments aren't as well modeled or understood in iceberg dynamics

      ☞ ViG *might* better capture those relationships



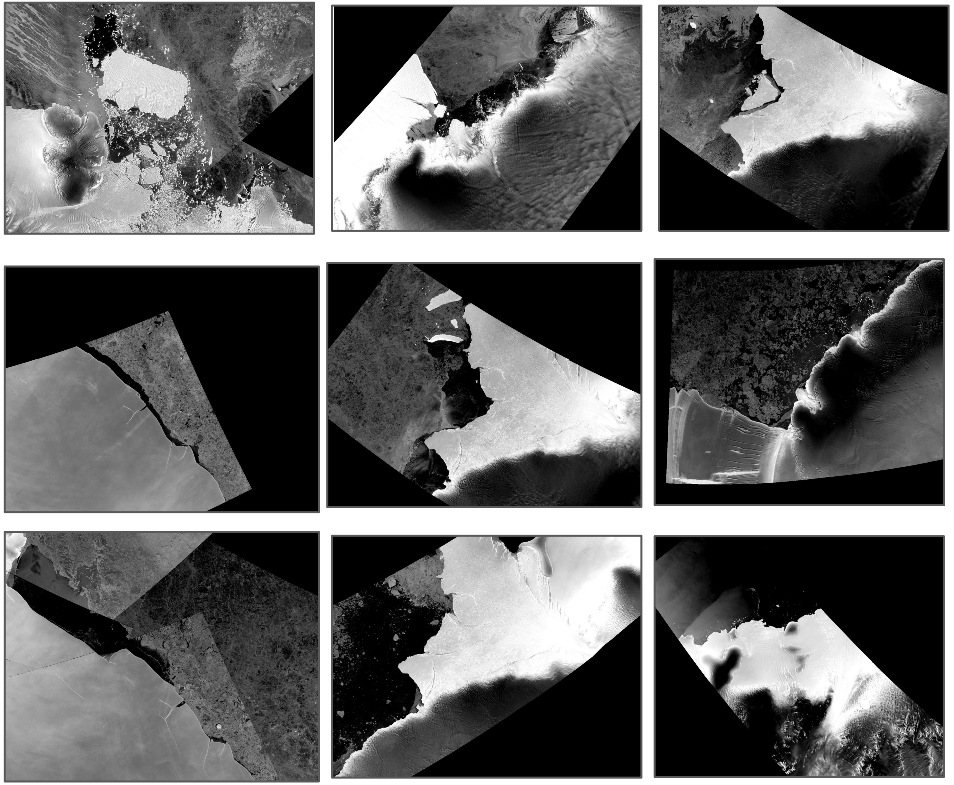


Reference: *Vision GNN: An Image is Worth Graph of Nodes*, *CS231n: Deep Learning for Computer Vision Lecture 1: Introduction*

Dataset

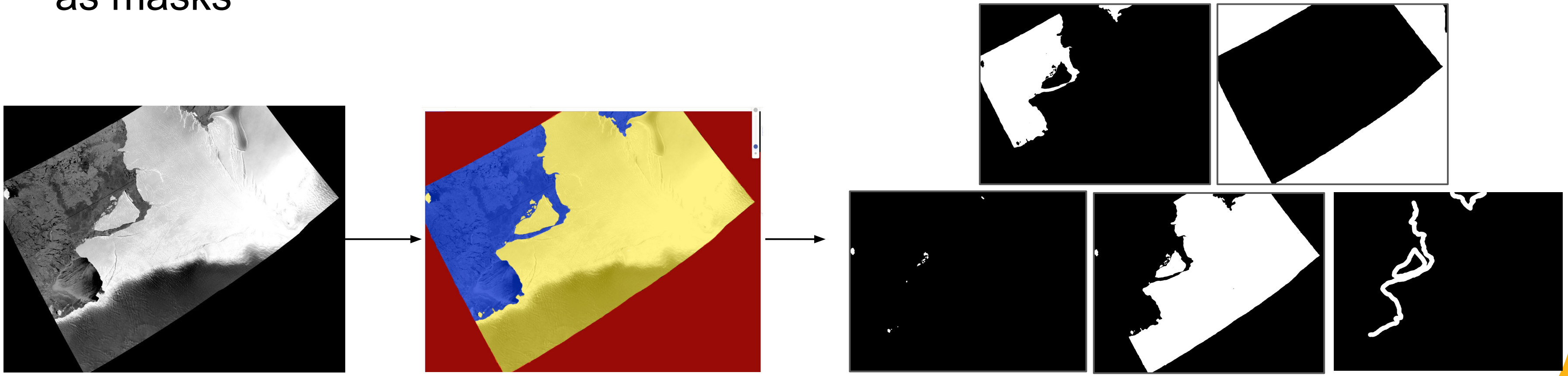
Dataset

- 90 unlabeled images of SAR imagery from Copernicus (Antarctic region)
- Each image contains ice, ocean, fragments, boundary, no data regions in different complex patterns.



Labeling Process: Label Studio

- Since the experiment later on requires annotated data as ground-truth, we had to manually annotate some of them using Label Studio.
- The procedure includes manually segmenting the class and exporting them as masks

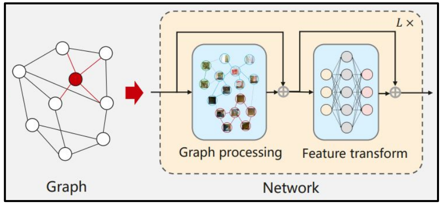


Approach

Our Experiment: Comparison between 2 strategies

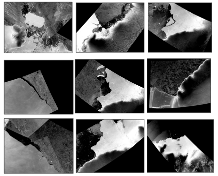
Baseline: ImageNet-pretrained ViG Fine-tune on labeled SAR iceberg data

ImageNet Dataset

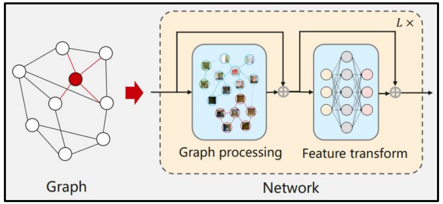


ViG

Finetune



Labeled
SAR Iceberg
Dataset



Finetuned ViG

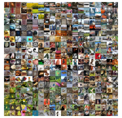
Source Domain
ImageNet Dataset → SAR iceberg imagery
Target Domain

Reference: https://cs.stanford.edu/people/karpathy/cnnembed/, Vision GNN: An Image is Worth Graph of Nodes

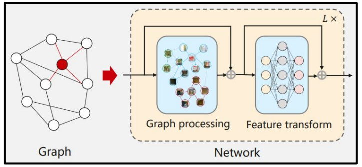
Our Experiment: Comparison between 2 strategies

Deluxe: ImageNet-pretrained ViG + Self-Supervised Representation learning

+ Fine-tune on labeled SAR iceberg data

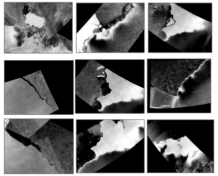


ImageNet
Dataset

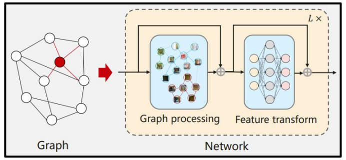


ViG

Self-Supervised

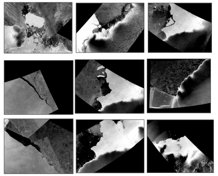


Unlabeled
Reconstruction to improve representations

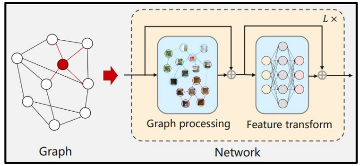


Adapted ViG

Finetune



Labeled
Classification Task



Adapted + Fintuned ViG

Source Domain | Target Domain

ImageNet Dataset → SAR iceberg imagery

Reference: https://cs.stanford.edu/people/karpathy/cnnembed/, *Vision GNN: An Image is Worth Graph of Nodes*

Task Overview

Architecture Evaluation

- Assessing **Vision Graph Neural Network (ViG)** as a viable framework for patch-level SAR classification

Comparing Two Strategies

- **Baseline:** Finetuning
- **Deluxe:** Self-Supervised Learning + Finetuning

Reference:

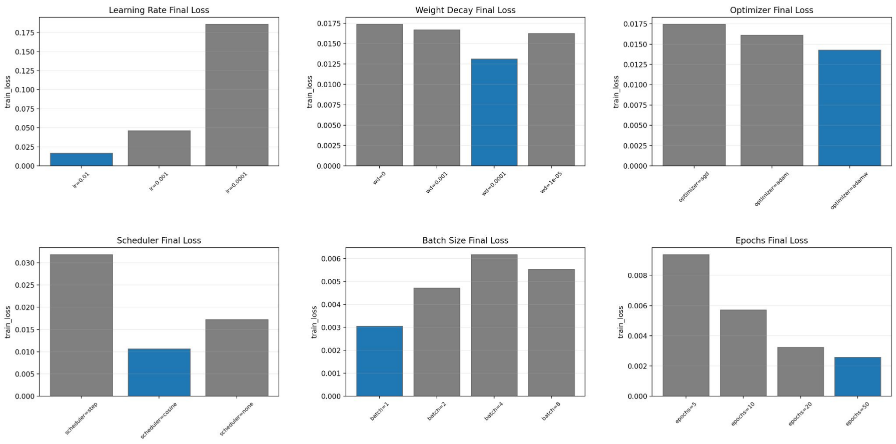
Training Dynamics and Monitoring

Hyperparameters

- **Setup**: One hyperparameter group at a time while keeping the rest of the setup fixed
- **Goal**: Identify the hyperparameter settings that
produce the lowest training loss
- **Hyperparameters (In-Order)**: learning rate, weight decay, optimizer, scheduler, batch size, and number of epochs.

Best Hyperparameters for each Learning step

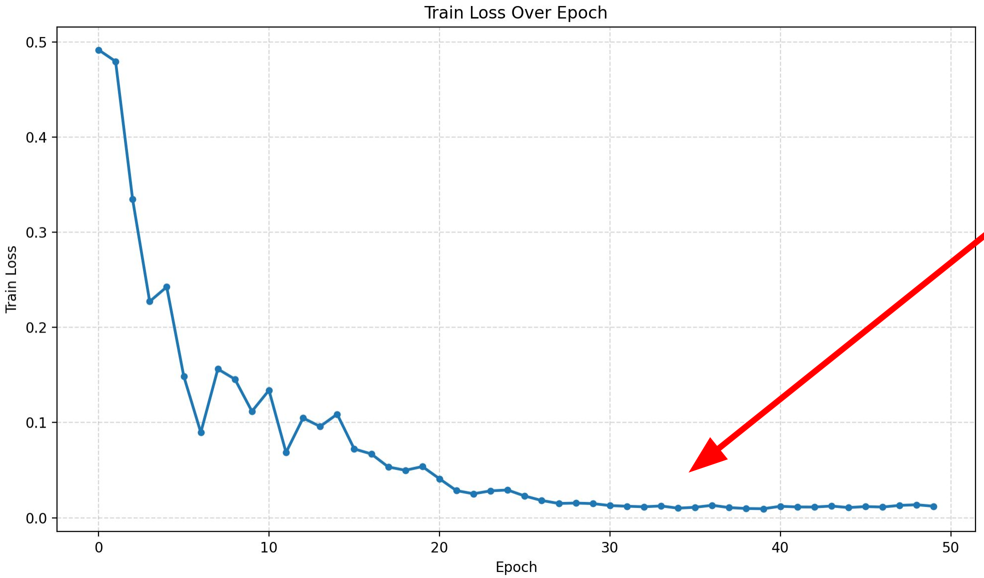
	Baseline Finetuning	Self-Supervised (Reconstruction)	Deluxe Finetuning
Learning Rate	0.001	0.01	0.001
Weight Decay	0.001	0.001	0.001
Optimizer	Adamw	Adamw	Adamw
Scheduler	None	Cosine	Cosine
Batch Size	1	1	2
Epoch	Higher Wins	Higher Wins	Higher Wins



Blue marks the lowest final train_loss in each sweep. Lower train_loss is better.

Monitoring via Training Loss

Baseline Finetuning Monitoring Illustration

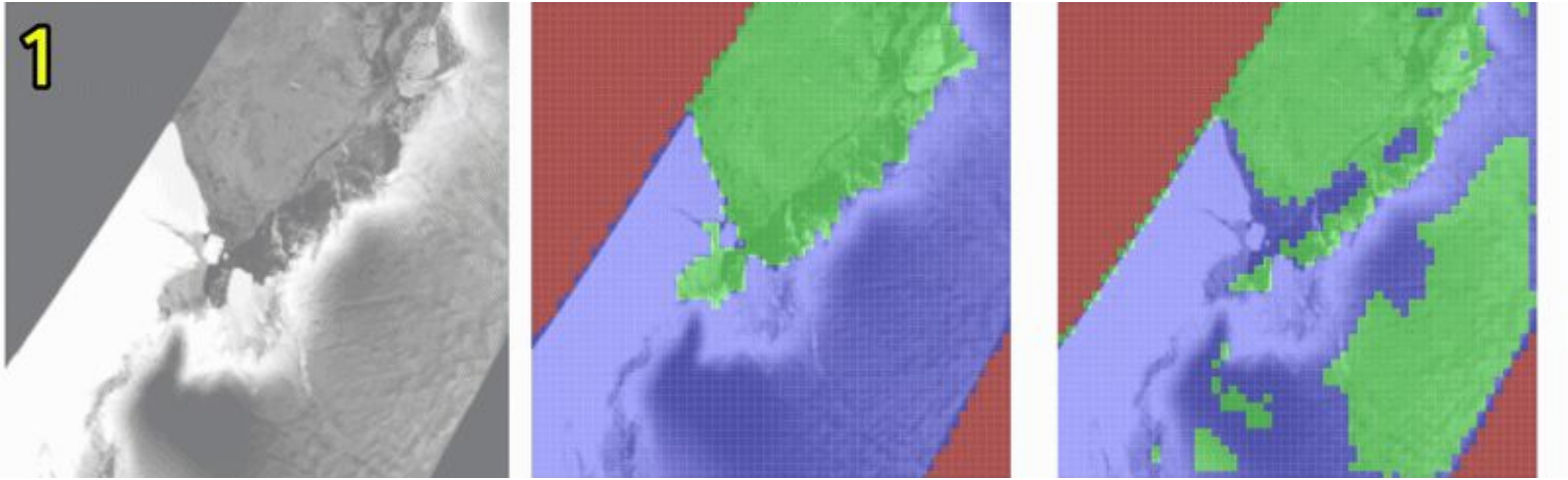


We could probably stop training here

We monitor the model convergence by plotting loss values against training iterations/epochs. Visualization helps diagnose model performance more <u>easily</u>

Monitoring via Training Predicted Output

Original Image | Ground Truth Overlay | Predicted Output During Training



1 = Epoch

= No Data

= Ocean

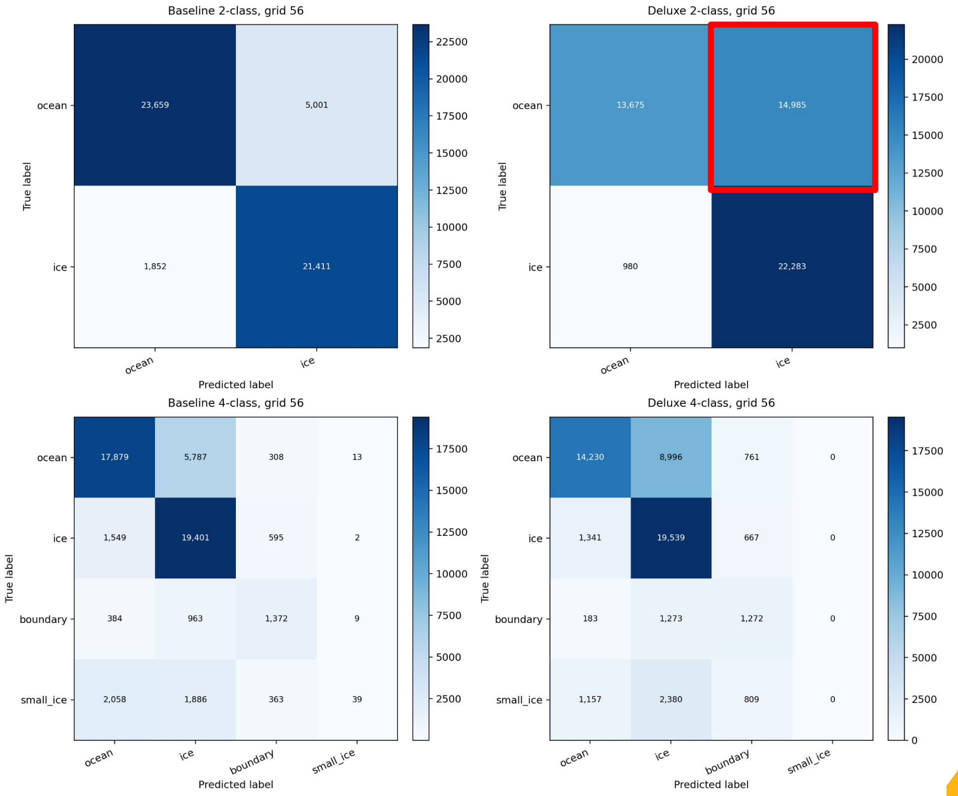
= Ice

Inference and Evaluation

Confusion Matrix Grid 56

- Baseline model is clearly stronger
  - Baseline correctly predicts more while making fewer mistakes
- Deluxe model struggles
  - It mistakes many ocean patches as ice

Remember This Note

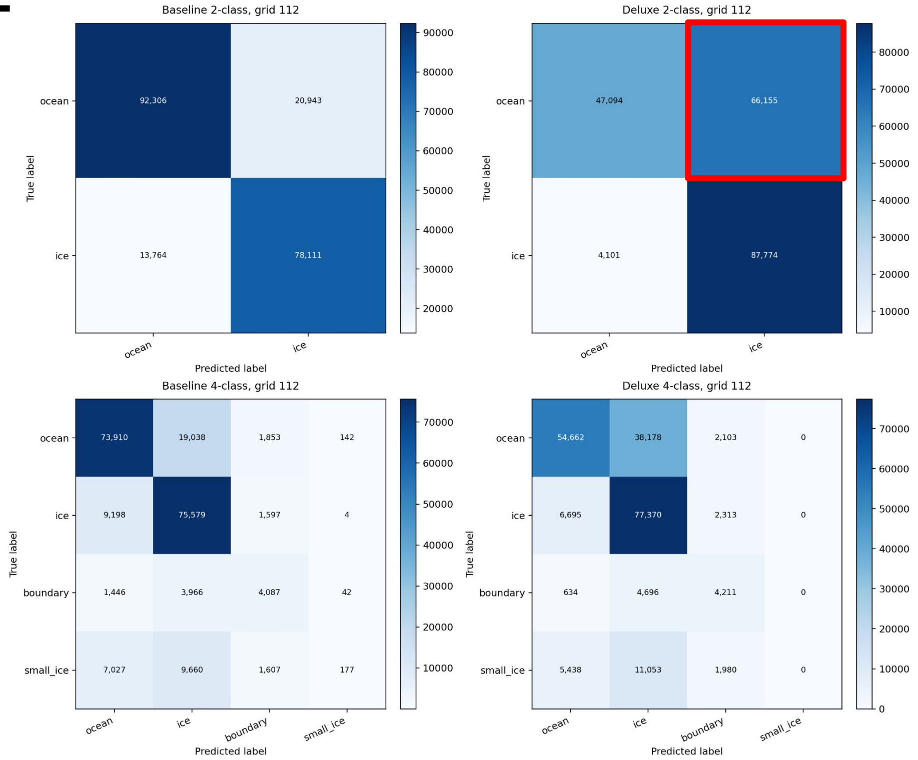


Confusion Matrix Grid 112

- Baseline model is clearly stronger
  - Baseline correctly predicts more while making fewer mistakes

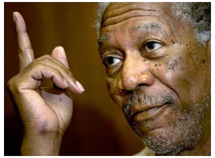
- Deluxe model struggles
  - It mistakes many ocean patches as ice

Remember This Note

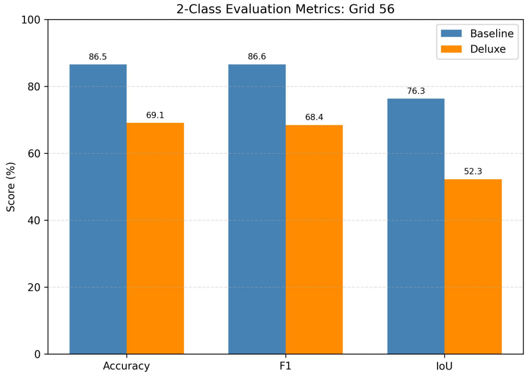


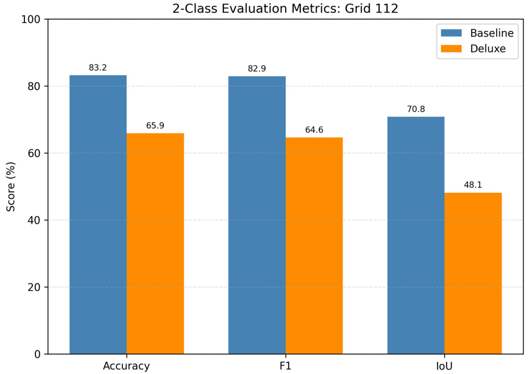
Classification Evaluation Metrics

- Accuracy, F1, and Intersection over Union (Coverage)
  - Baseline performs <u>significantly</u> better
- Best performance shown in Grid 56
  - Larger grid is better
- Lower IoU scores in both cases suggest
  - The models may detect the correct general region, but the boundaries/details are less accurate



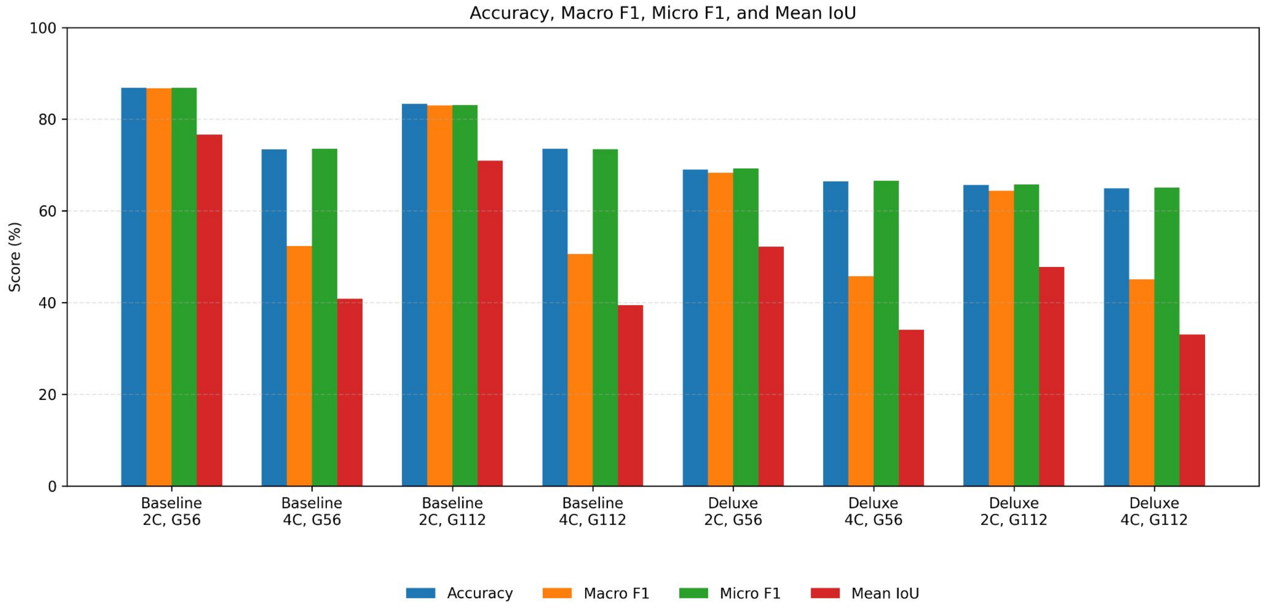
Remember This Note





Classification Evaluation Metrics

- Comparing the evaluation metrics across all experiments we see that the Baseline outperformed the Deluxe in all cases

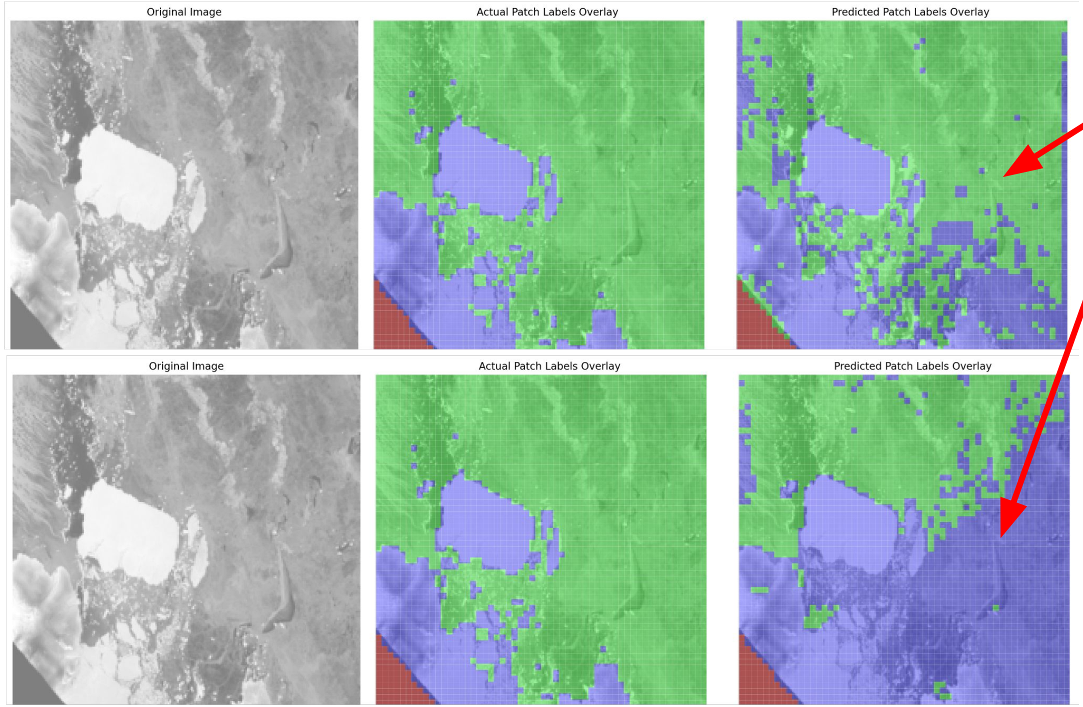


Compare Actual Image Output (Grid 56)

Original Image | Ground Truth Overlay | Predicted Output

Baseline

Deluxe



Compared to Baseline, the Deluxe model struggles with predicting ocean patches suggested by the Confusion Matrix table

= No Data

= Ocean

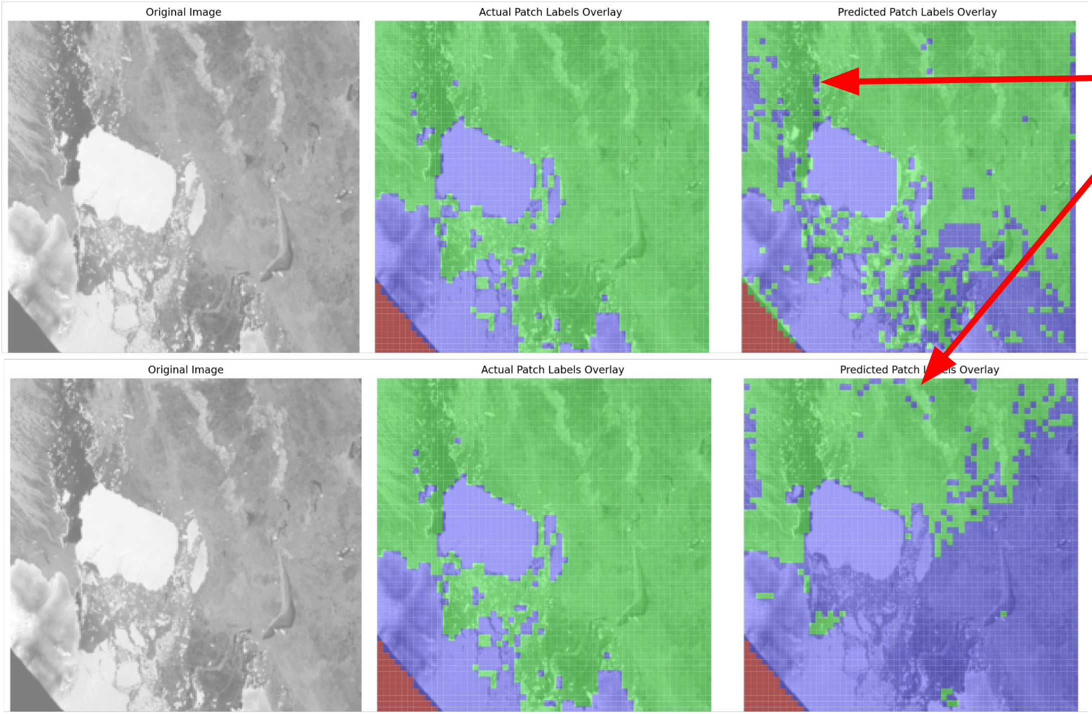
= Ice

Compare Actual Image Output (Grid 56)

Original Image | Ground Truth Overlay | Predicted Output

Baseline

Deluxe



Boundaries/details are less accurate suggested by the lower IoU metrics

= No Data

= Ocean

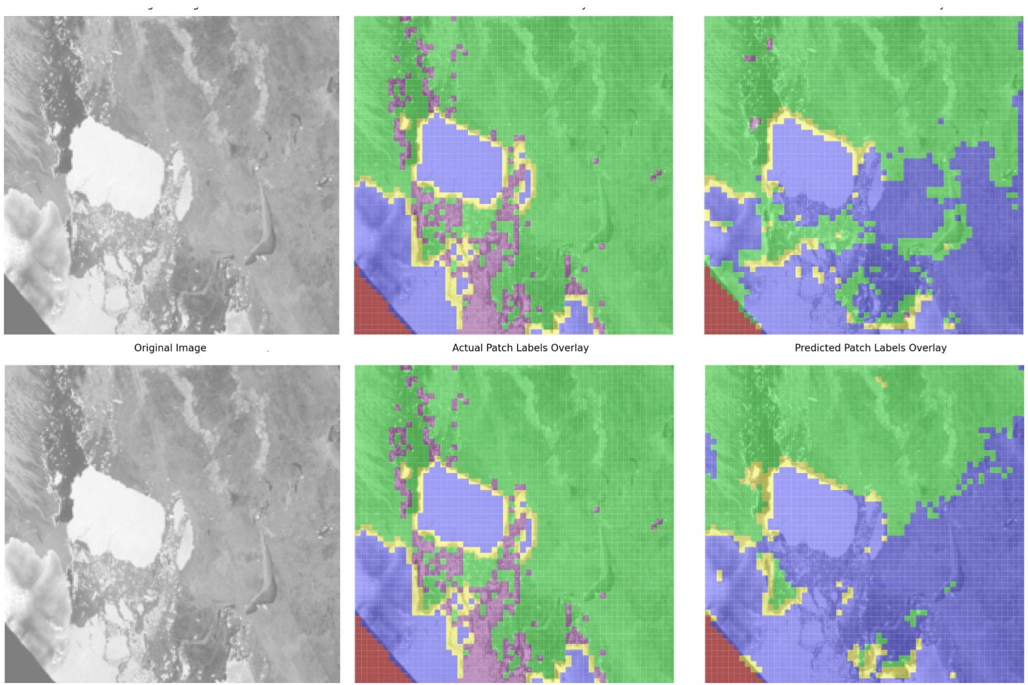
= Ice

Compare Actual Image Output (Grid 56)

Original Image | Ground Truth Overlay | Predicted Output

Baseline

Deluxe



Deluxe does not predict any small ice fragments compared to baseline, and baseline better predicts boundaries

= Small Ice Fragments

= Boundary

= No Data

= Ocean

= Ice

UMBC

Conclusion

Architecture Evaluation

- ViG backbone adaptation is viable (86.5% accuracy in baseline), though there are still many challenges.
- Larger grid size (56) is better at this stage.

Comparing Two Strategies

- Baseline (Direct Finetuning) is the stronger at this stage.
- Deluxe works in theory but might not in practice.
  Our hypothesis: Self-Supervised (Reconstruction) should help preserve SAR texture, but the objective might have not well aligned with the classification downstream task.

Limitations & Future Work

- Lack of diversity in both labeled/unlabeled data hurts the models
  - Data Augmentation (crop, resize, rotate) could help diversify our dataset
  - Add more diverse scenes of icebergs
- Annotation is time-consuming
  - Use current pipeline to speed up manual annotation
  - Automate ice/ocean annotation
- Ground truth labels may contain noise or bias when manually annotating
  - Set up stronger guidelines
- Model performance is sensitive to grid size
  - Test more patch sizes or overlapping patches

Questions?