MUSA 650 Final Report
Singapore Ship Detection

Eugene Chong
Madhura Gurav
John Michael LaSalle

Introduction

The Problem: Can we classify ships in Sentinel-2 imagery using a model trained only on imagery from a different satellite with a different spatial resolution?

Why it matters: The proliferation of satellite imagery providers creates more opportunities for researchers but labelled data may not be available for every provider. If models can be trained on existing labelled data and applied across data from different providers, researchers would have significantly more resources to work with.


Data

Train - Shipsnet

- Created by Bob Hammell
- Planet PlanetScope imagery
- ~3 meter resolution
- 4,000 labelled patches
  - 1,000 Ships
  - 3,000 not ships
- 80px x 80px x 3 bands
- Classifies partial ships as not-ship
- Accessed via Kaggle API

Test - Sentinel-2

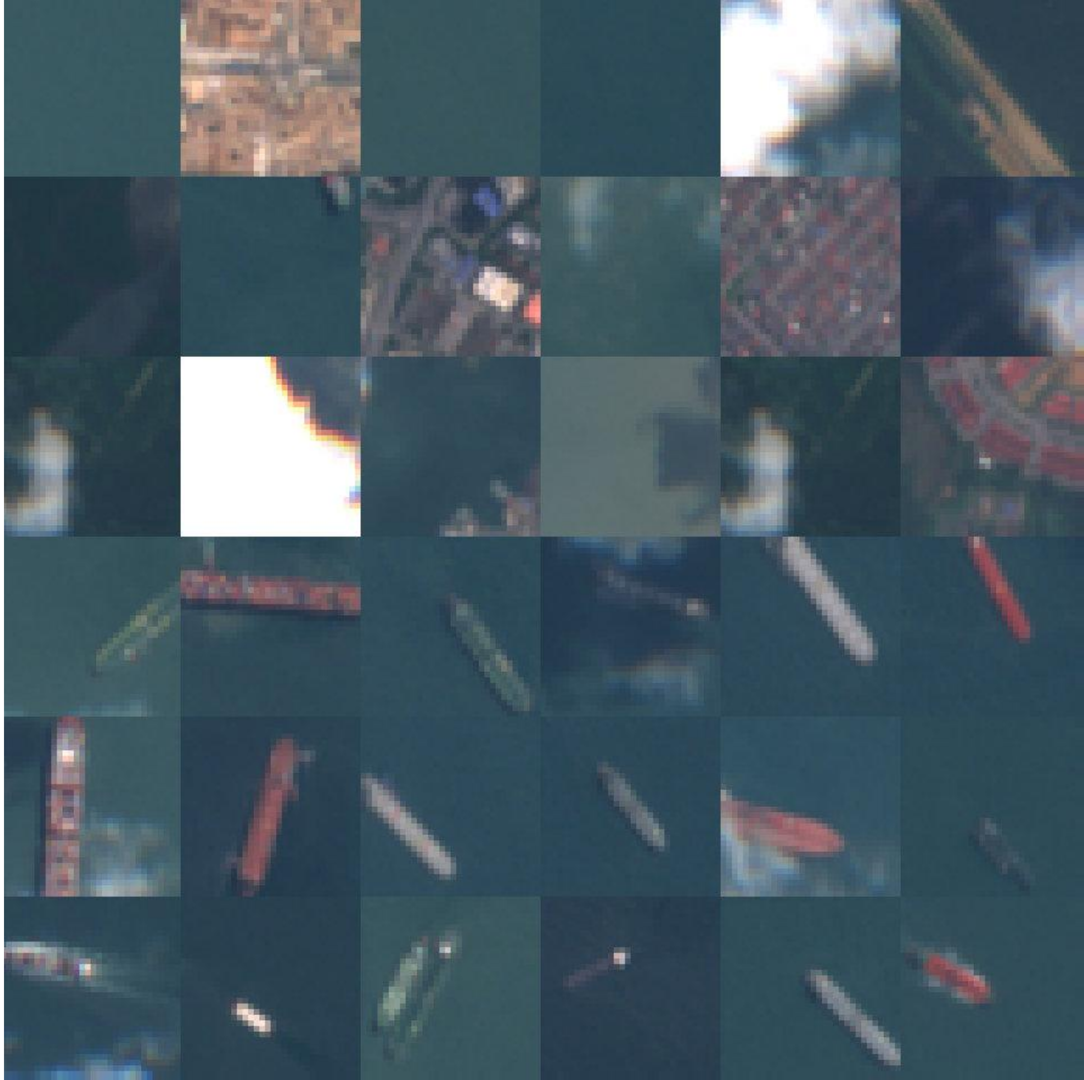
- European Space Agency
- 10 meter resolution
- 100 km x 100 km x 13 bands
- Unlabelled
- Accessed via Copernicus Open Access Hub API

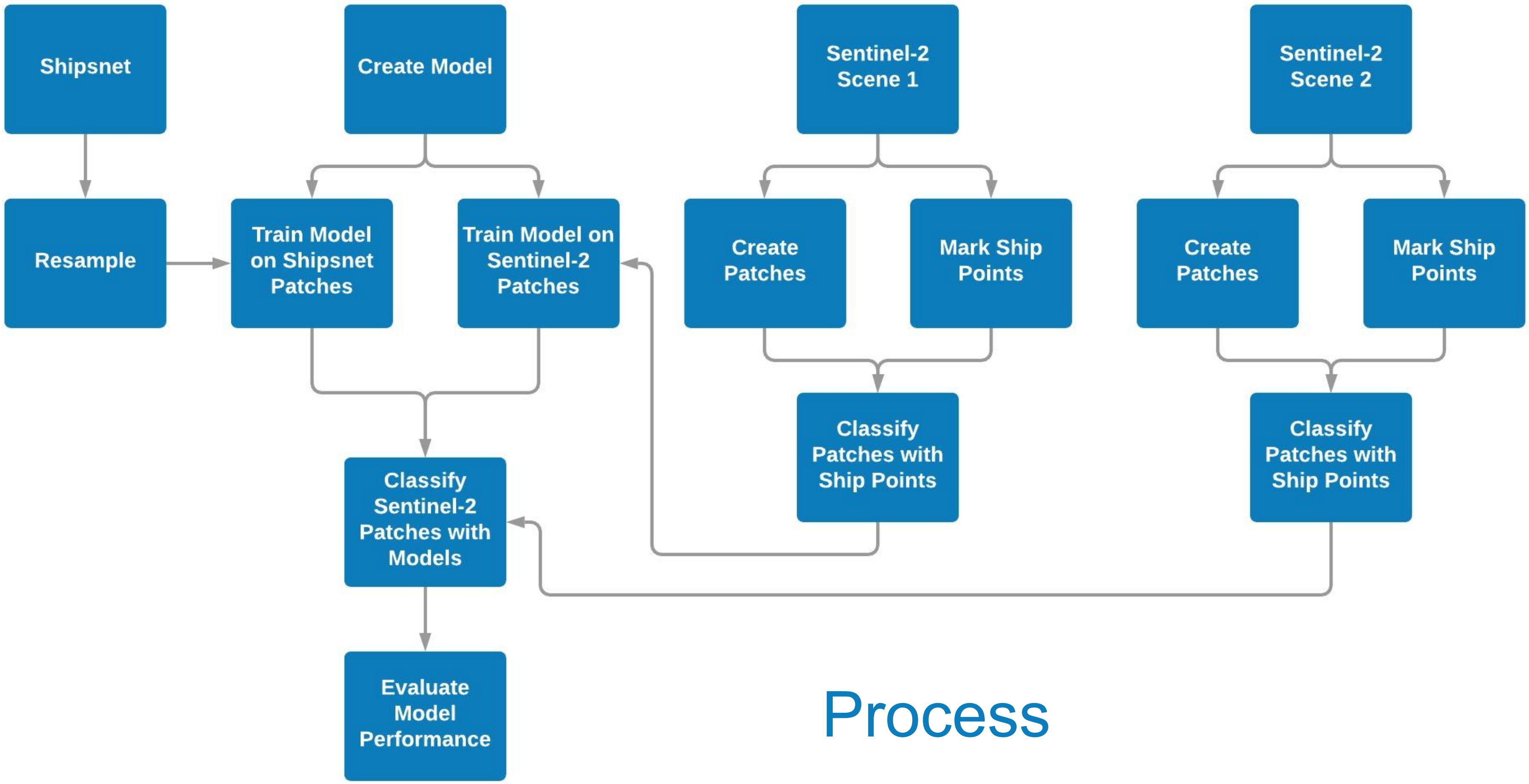
Methods

Process

- Train a model on labeled Shipsnet data
- Evaluate the model on Sentinel data, which we label ourselves
- As benchmarks, train models with the same parameters on a subset of Sentinel data and evaluate on unseen Sentinel data

Our goal is to understand how reliably a model trained on labelled data can be used to evaluate data from a different source with a different resolution.

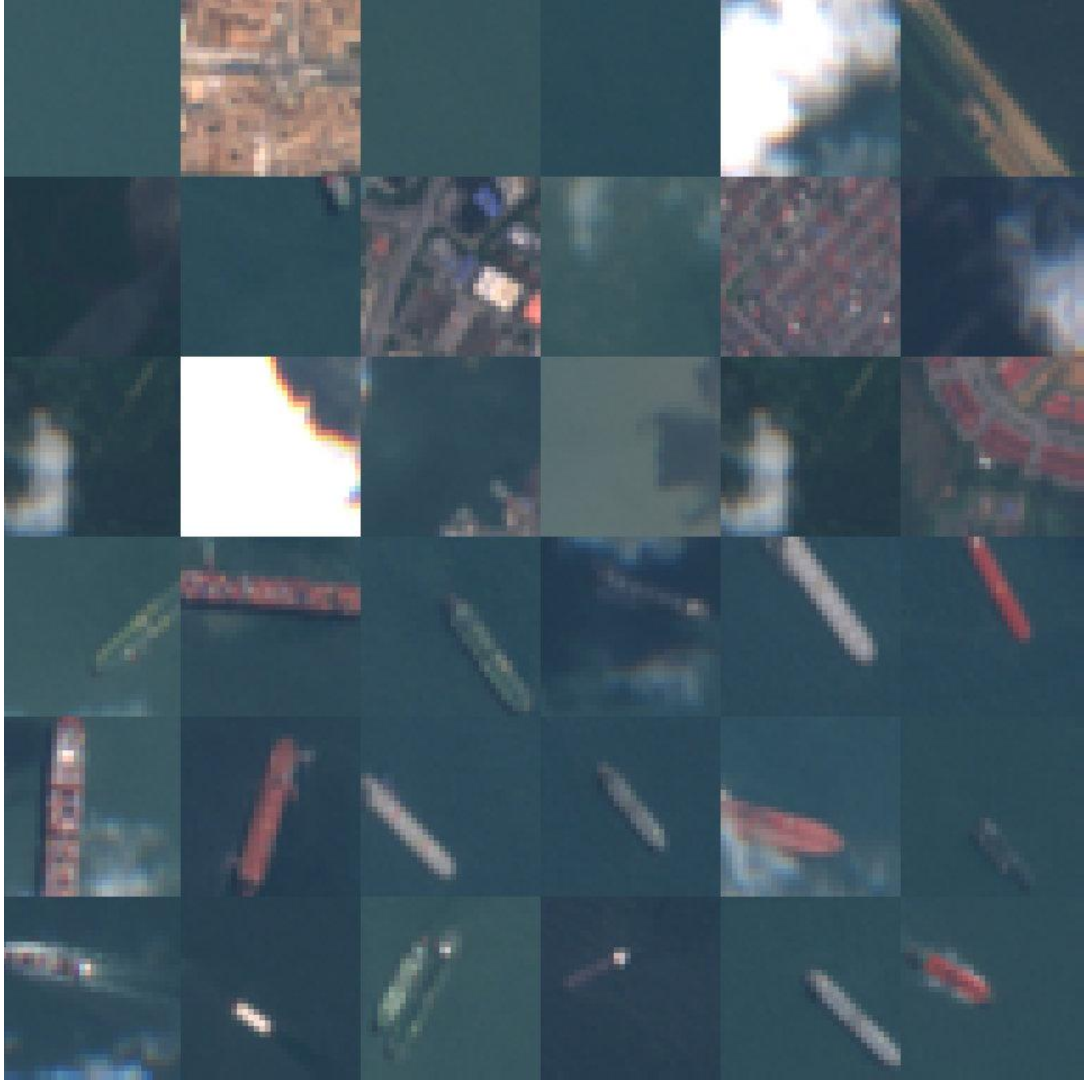



Shipsnet
Create Model
Sentinel-2 Scene 1
Sentinel-2 Scene 2
Resample
Train Model on Shipsnet Patches
Train Model on Sentinel-2 Patches
Create Patches
Mark Ship Points
Create Patches
Mark Ship Points
Classify Patches with Ship Points
Classify Patches with Ship Points
Classify Sentinel-2 Patches with Models
Evaluate Model Performance

Process

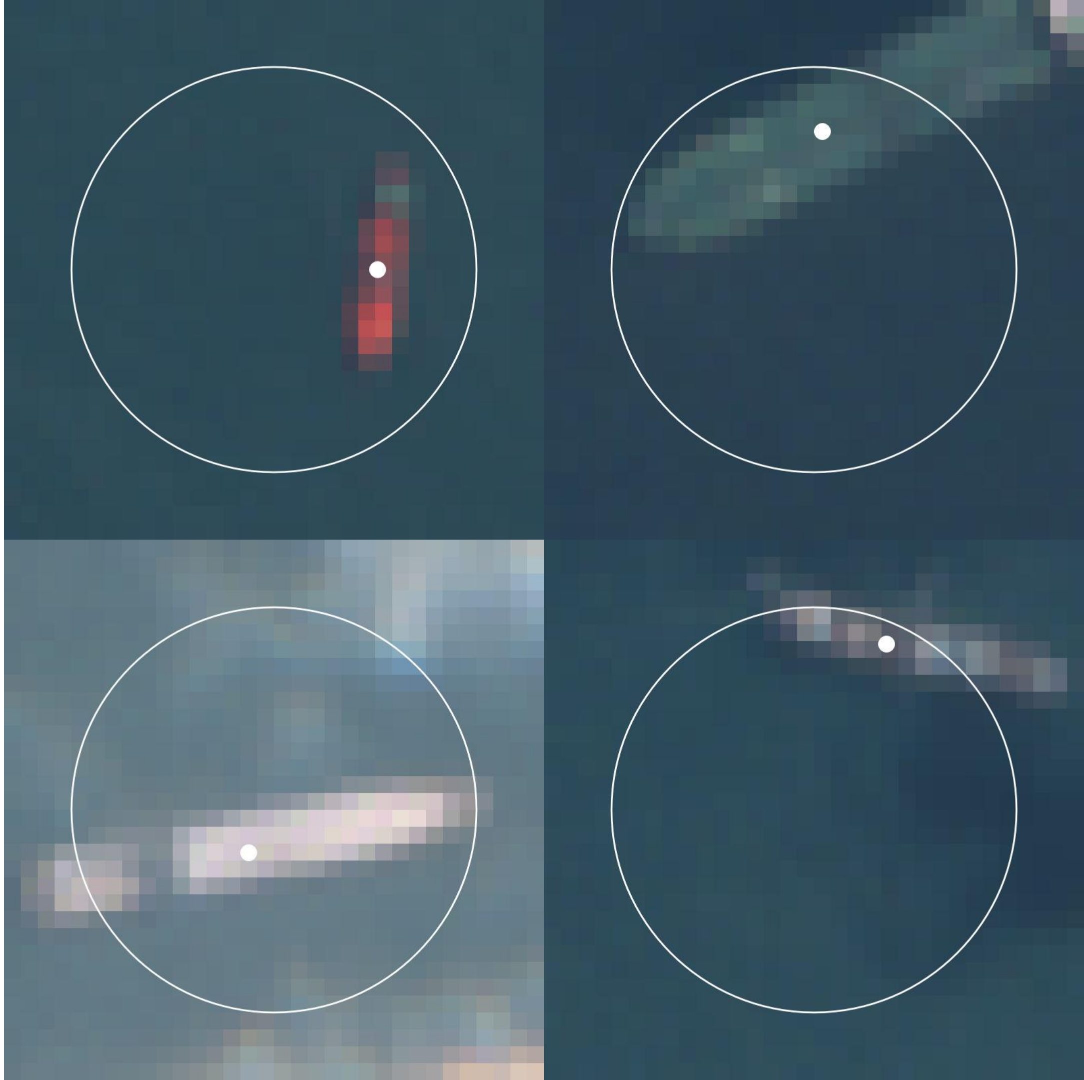
Sentinel Patches

- 32px by 32px with 3 bands
- 16px overlap
- About 470,000 patches per image across two images



Ship Point Intersection

- We manually marked ship locations in QGIS. About 650 ships total across two scenes, so the vast majority of patches are non-ship.
- For each patch, we marked as “ship” if the ship location was within some distance from the center.



Resampling

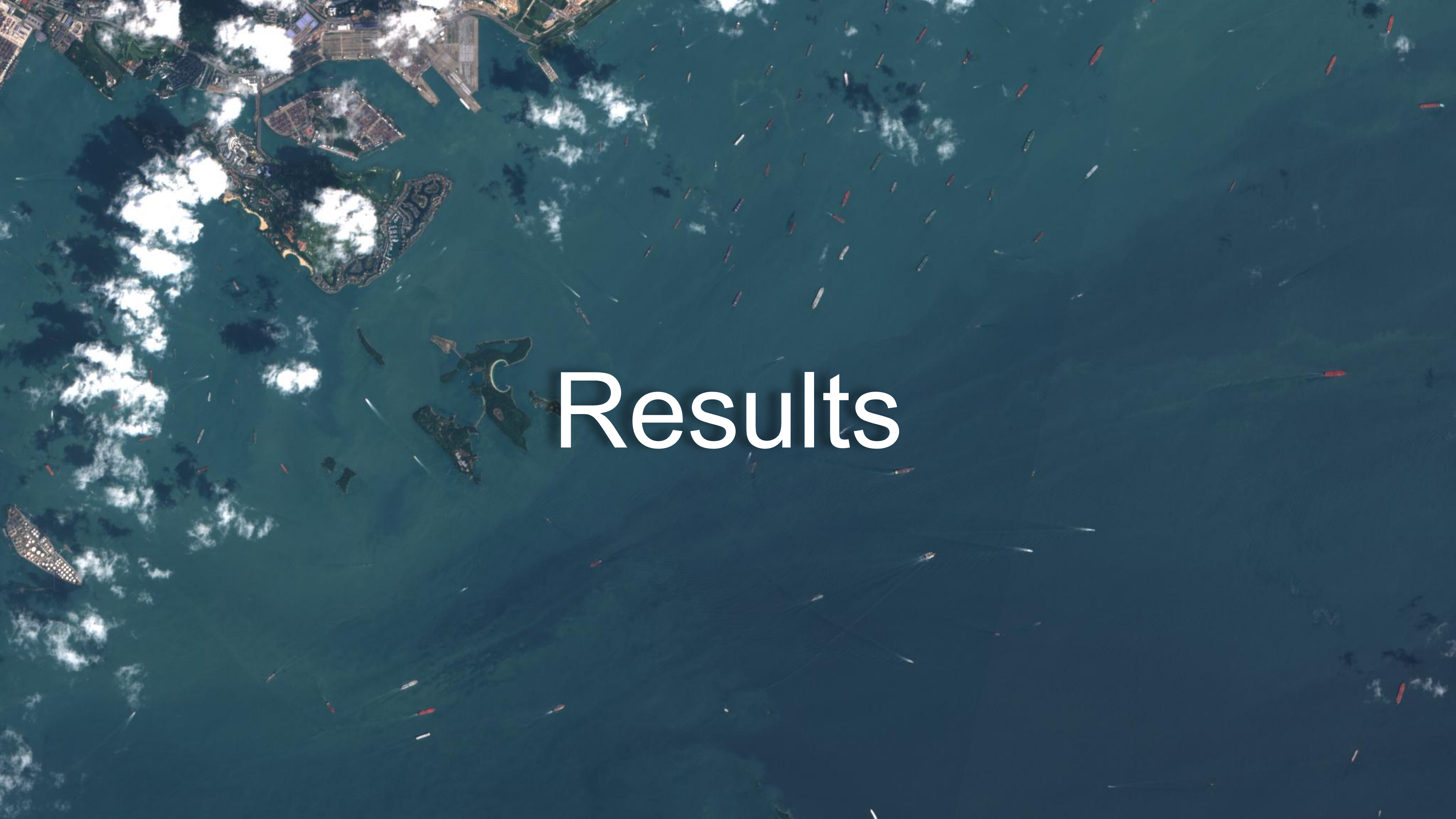
- Downsampled Shipsnet patches to match the Sentinel resolution (from ~3m to 10m)
- Above: original images
- Below: resampled images

Model Training

We trained 4 types of models on a 50% training set of the pre-labelled, downsampled Shipsnet data and evaluated on the remaining data. Interestingly, the shallow CNN had the best performance, and the deep CNN that used a pre-trained model as a convolutional base had significantly worse performance. The deeper models tended to overfit the training data, outweighing any benefits gained from learning more lower level features.

Model	**Description** All models trained on a 50% subset of the downsampled Shipsnet data	**Accuracy on Shipsnet Data Test Set**
1. SVM	• Trained on flattened RGB vectors. • Hyperparameters optimized using nested cross-validation and GridSearchCV	Average of 94.5% across folds
2. Shallow CNN	• Trained on 32x32x3 tensors • Only one hidden convolutional layer • Some data augmentation (shearing, zooming, flipping)	97.4%
3. Deep CNN	• Several convolutional/pooling layers and a hidden dense layer • EarlyStopping and ReduceLRonPlateau callbacks • More aggressive data augmentation (shifting, rotation, more shearing and zooming)	96.2%
4. CNN with Transfer Learning	• Tried several pre-trained models for the convolutional base. DenseNet121 had best performance • Overall, performance was much worse than simpler models	78.7%

Results



Model Results (1 of 2)

For each model, we present two sets of results. The **Shipsnet** results are for models trained on the Shipsnet data (as described on the previous slide) and evaluated on the Sentinel-2 data. The **Baseline** results are for models trained on a subset of the Sentinel-2 data and evaluated on the remaining Sentinel-2 data. This is the traditional approach to training supervised machine learning models on satellite images, but it required the expensive process of labelling the Sentinel-2 data ourselves, which may not be feasible for more complicated classification tasks.

The Baseline versions of the models identified significantly more positive cases in the Sentinel-2 patches than the Shipsnet versions were able to. We weren’t surprised to see that the Baselines performed significantly better, but we were surprised to see how poorly the Shipsnet versions transferred to Sentinel-2. The Baseline models did have much higher rates of false positives than the Shipsnet versions, shown by their lower specificities.

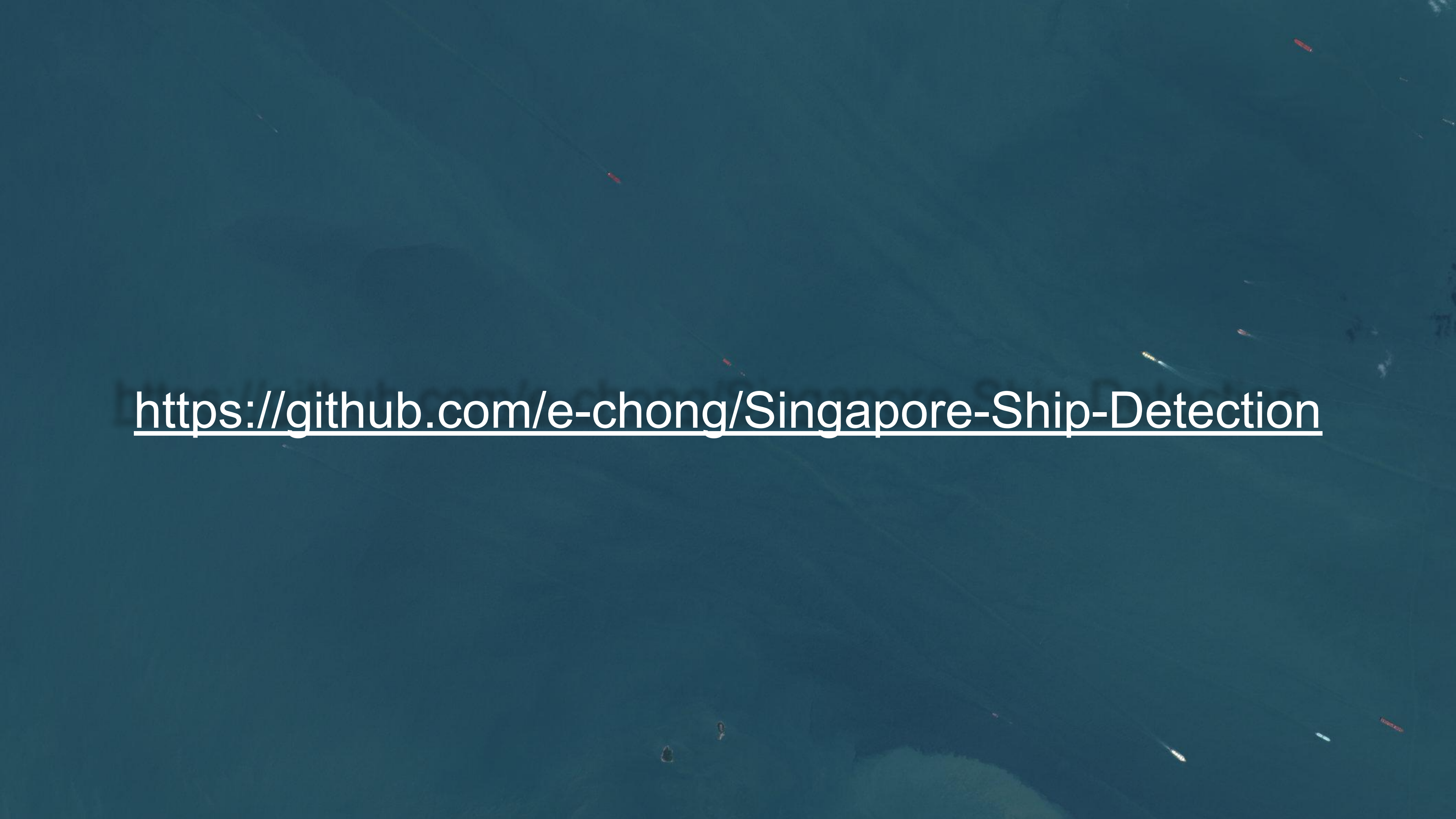
Model Results (2 of 2)

Model	Accuracy	Specificity	Sensitivity	Precision
1. SVM (Shipsnet)	0.9944	0.9955	0.0862	0.0230
1. SVM (Baseline)	0.9075	0.9076	0.8000	0.0082
2. Shallow CNN (Shipsnet)	0.9994	0.9994	0.0517	0.0938
2. Shallow CNN (Baseline)	0.9657	0.9658	0.8889	0.0243
3. Deep CNN (Shipsnet)	0.9964	0.9975	0.1379	0.0635
3. Deep CNN (Baseline)	0.9394	0.9394	0.9556	0.0149
4. CNN with Transfer Learning (Shipsnet)	0.8781	0.8788	0.2931	0.003
4. CNN with Transfer Learning (Baseline)	0.8562	0.8566	0.4222	0.0028

Discussion

Key Takeaways

- Overall, it is not easy to combine data sources.
- The Shipsnet data was not representative of all ships because it only included large cargo ships centered in the patch, limiting how effective the models could be even with data augmentation
- Downsampling images reduces model performance by obscuring features
- Using only RGB bands makes differentiating between water and forest difficult. Using NDWI and NDVI could have reduced the number of false positive and false negative results.
- Shallower models seemed to work better. Deeper models seemed to overfit the training data, and features learned in deeper models didn’t translate well to new data.
- For this classification task, researchers who want to or need to use Sentinel-2 imagery would probably be better off creating their own training data than using a model trained on pre-labelled data from a different source.
- It might help to combine the data from multiple Sentinel-2 scenes since the scenes tend to not have sufficient ship patches to train the baseline models.


https://github.com/e-chong/Singapore-Ship-Detection