Automated Road Network Extraction and Route Travel Time Estimation from Satellite Imagery

Erdős Institute Data Science Bootcamp Project

Team Member: Heyang Long, Jason Guo
Mentor: Diego Galdino

The Problem

Goal: Leveraging computer vision algorithm to detect road networks from satellite images of cities. The extracted networks and their speed limits allows estimation of travel time between locations.

(SpaceNet5 Challenge: https://spacenet.ai/sn5-challenge/)

Dataset Overview

Two cities satellite image:

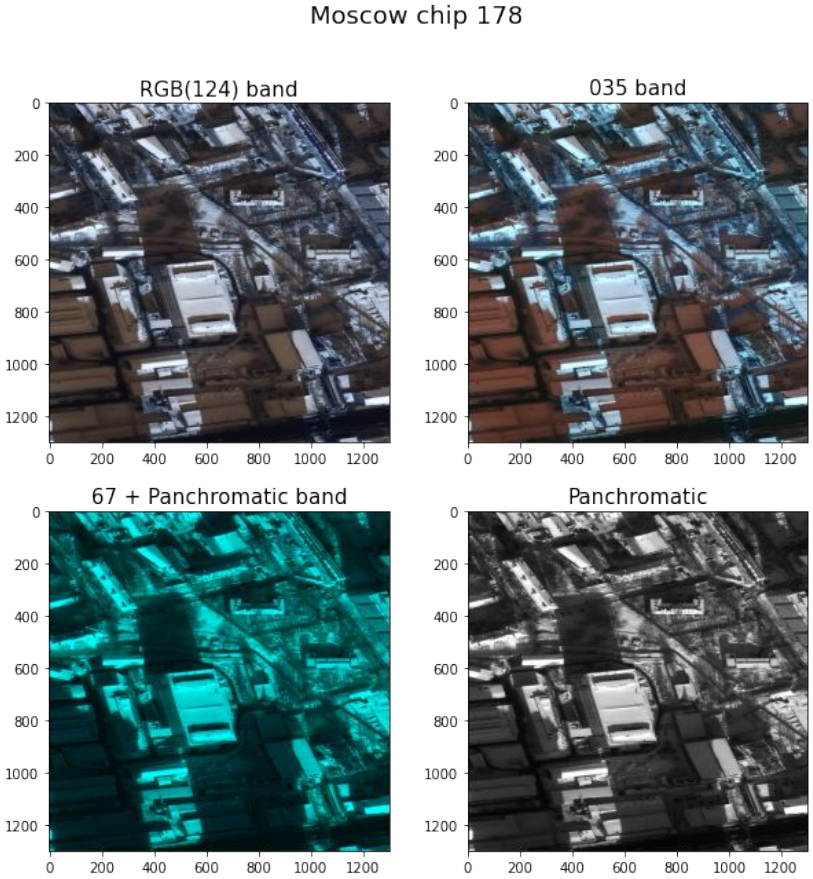
Moscow: total length: 2252.3 km, 1353 tiff files

Mumbai: total length: 1391.6 km, 1016 tiff files

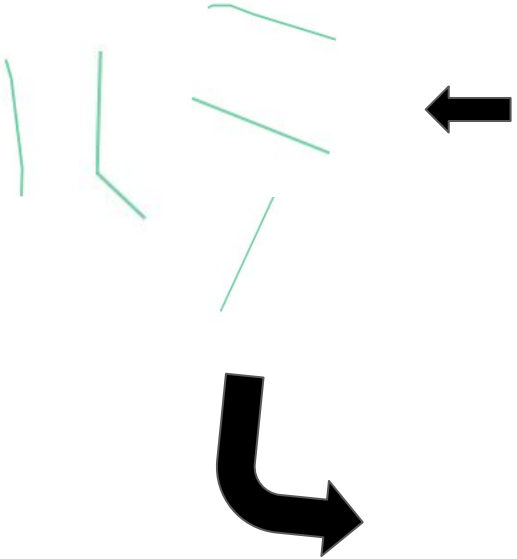
Tiff image files:

Pan-sharpenned multi-spectral images (1300*1300*8)
+ Panchromatic (1300*1300*1)

⇒ split into 4 png files each has 3 channels

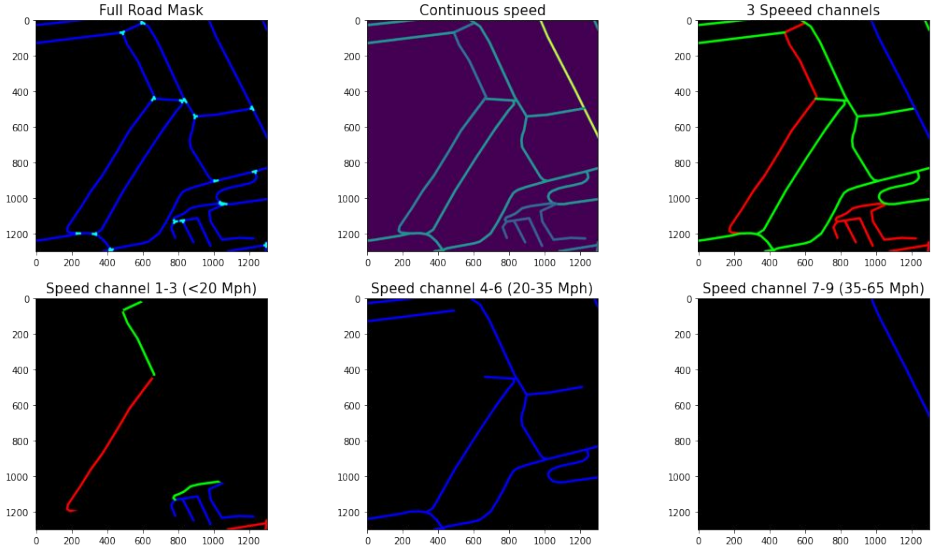


Data Preparation - Create Masks

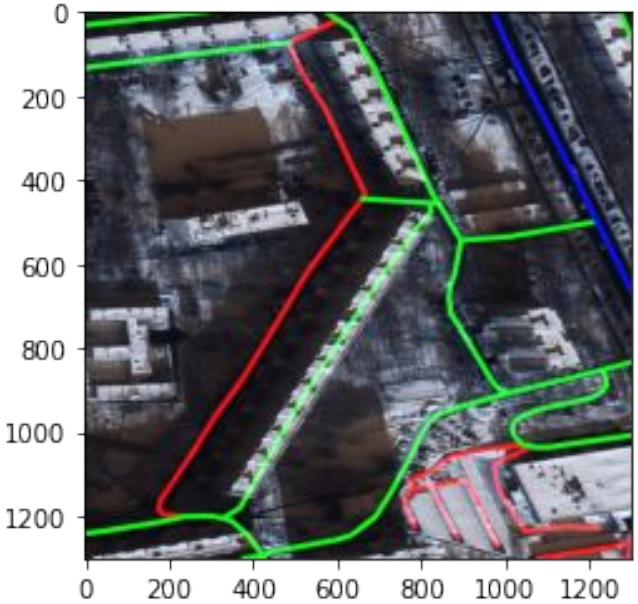


Road geometry and speed information in csv file

	ImageId	WKT_Pix	length_m	travel_time_s	speed
8593	SN5_roads_train_AOI_7_Moscow_chip1300	LINESTRING (1218.921328015625 499.127116709947...	59.492	2.957329	45.0
8594	SN5_roads_train_AOI_7_Moscow_chip1300	LINESTRING (1295.024837473407 1264.24377623200...	12.418	1.388914	20.0
8595	SN5_roads_train_AOI_7_Moscow_chip1300	LINESTRING (1078.604571962729 1299.99495349079...	44.062	4.928194	20.0
8596	SN5_roads_train_AOI_7_Moscow_chip1300	LINESTRING (1299.994953490794 1246.84674017503...	6.326	0.707543	20.0
8597	SN5_roads_train_AOI_7_Moscow_chip1300	LINESTRING (1062.504190951586 1033.84425793215...	47.759	4.273354	25.0



Data Preparation



- **Image augmentation**

  Shift, rotate, crop, resize, gaussian noise, contrast/ brightness/ saturation tune….

- **Data split**

  80% train, 10% validation, 10% test

Dataset Overview

Satellite images of different cities are sliced into 1300x1300 pixels chips (~1000 images per city). Original image data have multiple spectrum channels and are clipped into 3 RGB channels.

Road networks are manually labeled as . The The training dataset are manually labeled

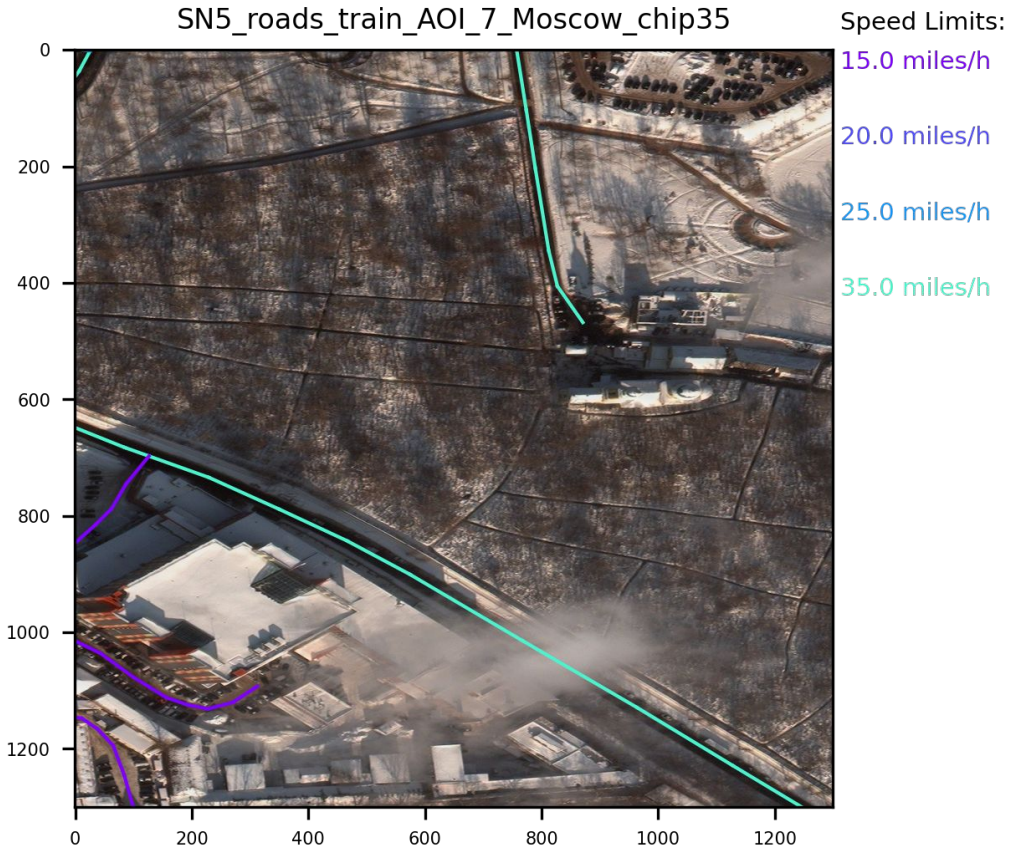


Image Segmentation Model

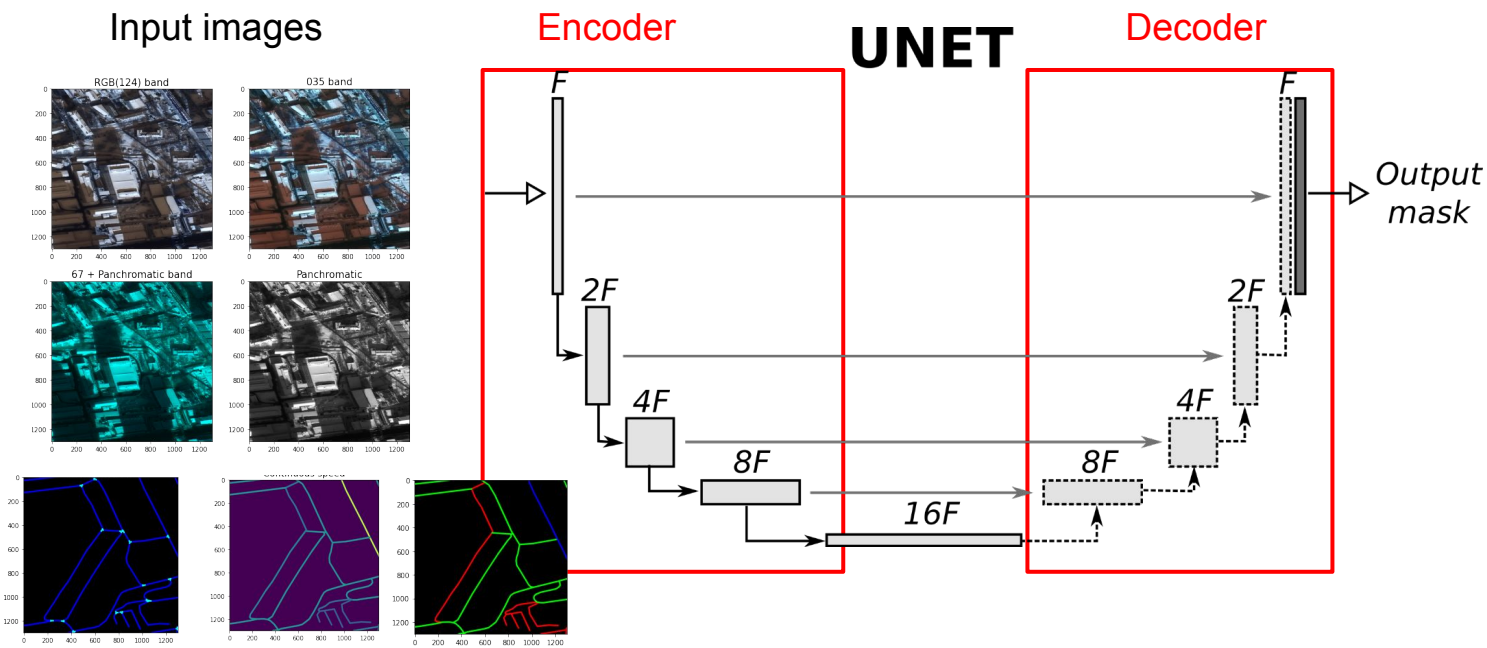


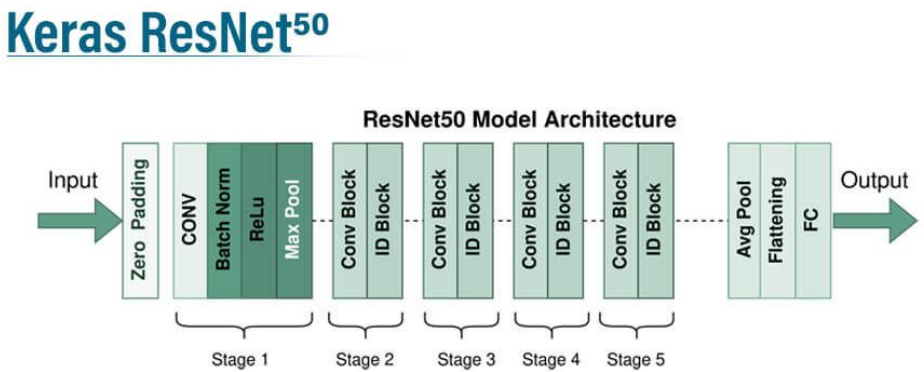
Image Segmentation Model

Final predictor ensembles 8 models (~90mins for each model training) :

- 4 folds pretrained ResNet50 encoder + Unet
- 4 folds pretrained SeResNeXt50 encoder + Unet

Combo loss with different weights:

- 1 Dice + 3 Focal for road mask
- Cross Entropy for speed mask
- MSE for continuous speed mask



Model training

The model inputs: 1300x1300x3 RNG images.

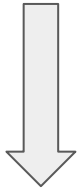
SpaceNet 5 dataset provides 1300x1300 image chips with corresponding road information files, which allows estimation of the speed limits of each road.

Loss function are defined as by:

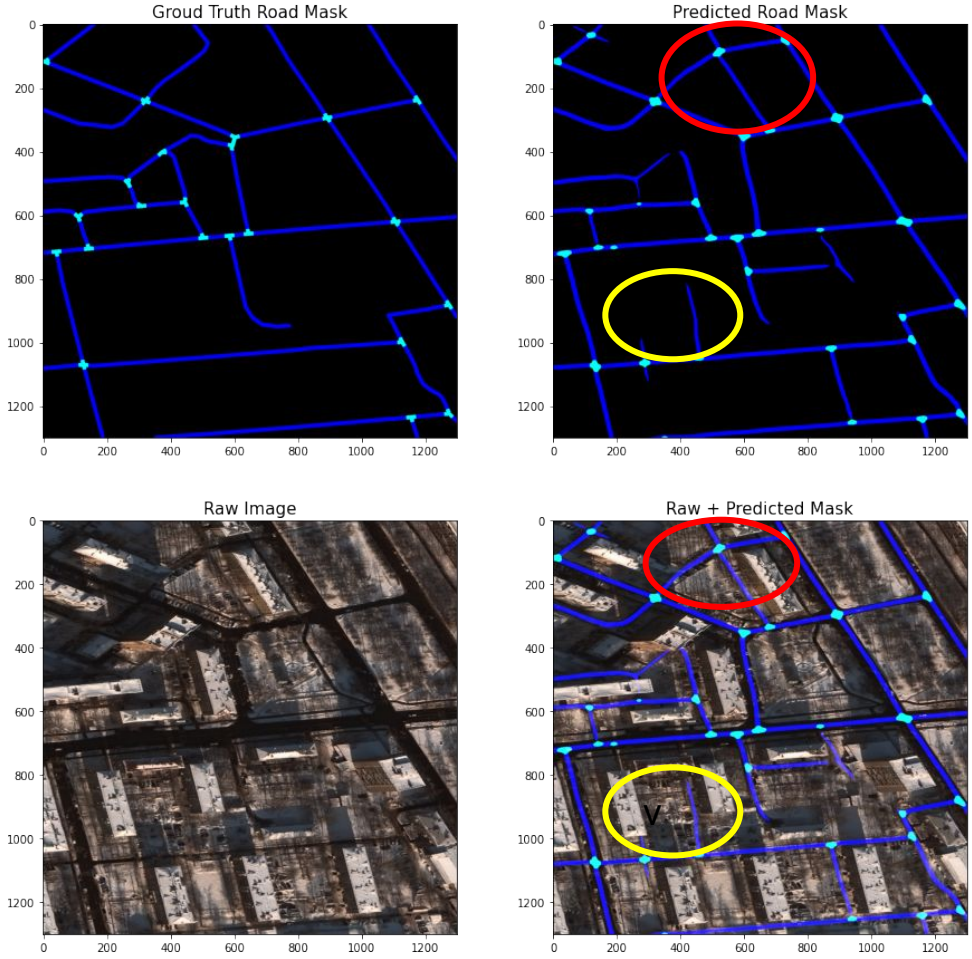
Prediction

- Road mask

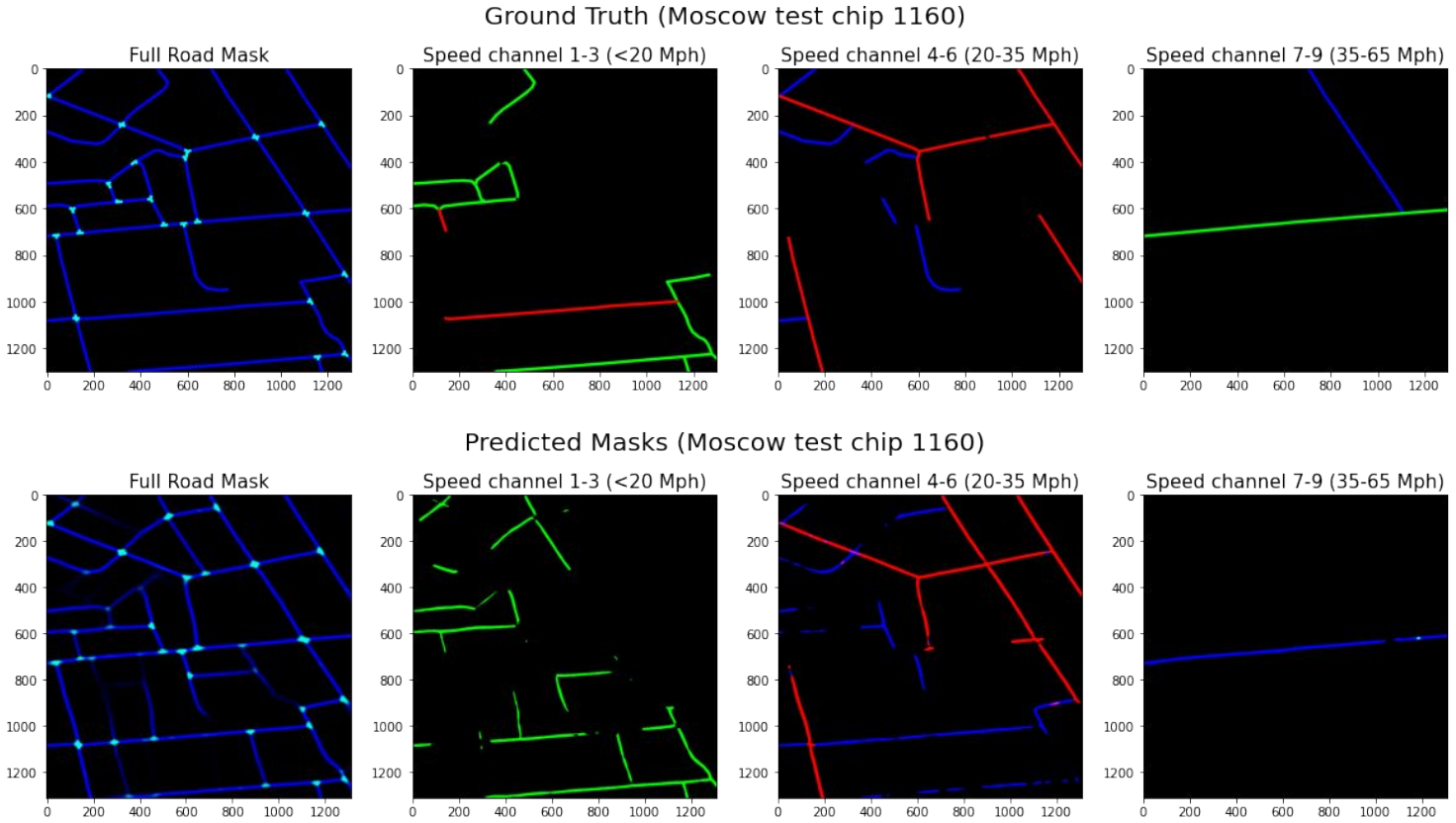
Overfitting or raw data flaws?



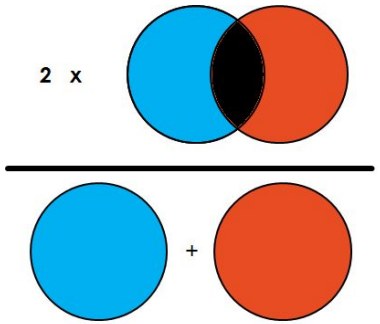
- Better annotated road masks are desired
- Adding penalty to loss function to prevent overfitting



Prediction


Ground Truth (Moscow test chip 1160)
Full Road Mask
Speed channel 1-3 (<20 Mph)
Speed channel 4-6 (20-35 Mph)
Speed channel 7-9 (35-65 Mph)
Predicted Masks (Moscow test chip 1160)
Full Road Mask
Speed channel 1-3 (<20 Mph)
Speed channel 4-6 (20-35 Mph)
Speed channel 7-9 (35-65 Mph)

Model Evaluation

Dice = 

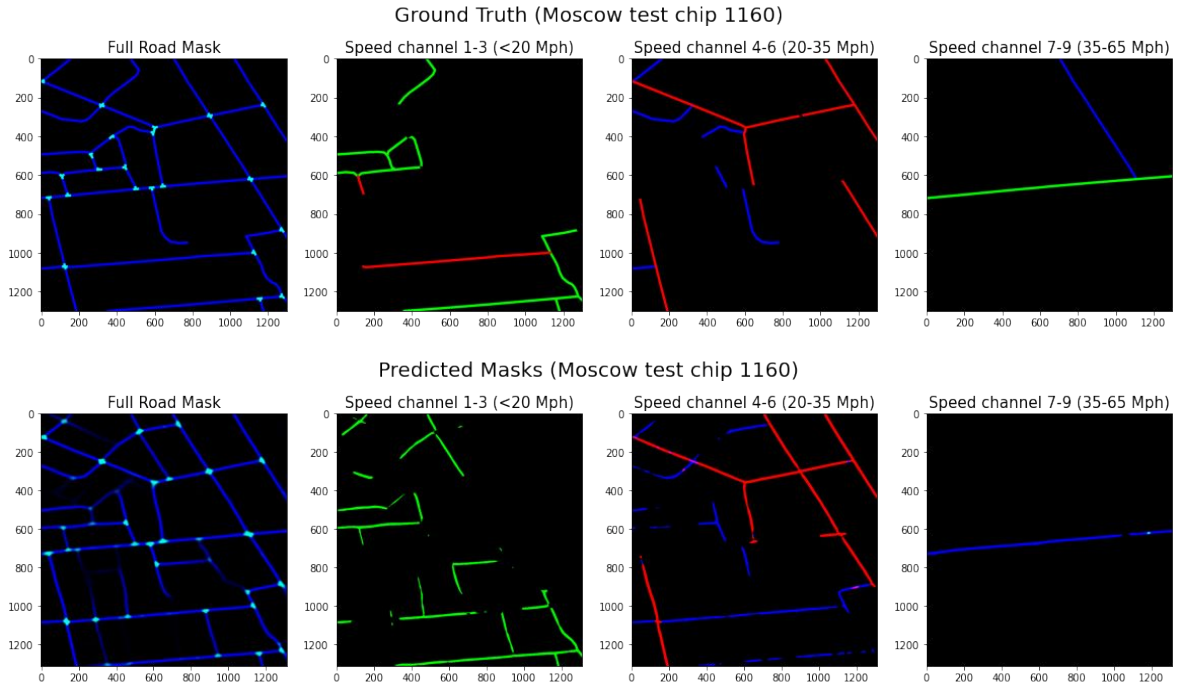
Dice score on test set:

Road mask: 0.6788,
Speed mask binned channel 1: 0.3926,
Speed mask binned channel 2 0.4639,
Speed mask binned channel 3: 0.6006

Recall we ensemble 8 deep learning models:

- 4 folds ResNet50 encoder + Unet (average Dice score ~0.65)
- 4 folds SeResNeXt50 encoder + Unet (average Dice score ~0.67)

Why speed channels prediction not good enough?



Dice score on test set:

Road mask: 0.6788,
Speed mask binned channel 1: 0.3926,
Speed mask binned channel 2 0.4639,
Speed mask binned channel 3: 0.6006

Combo loss with different weights:

1. 1 Dice + 3 Focal for road mask
2. Cross Entropy for speed mask
3. MSE for continuous speed mask

Carefully looking at the Combo loss, we think the weights added to speed mask's losses are way smaller than the road mask's loss, which could explain the performance.
Solution: tune the loss weights to get better performance for speed prediction

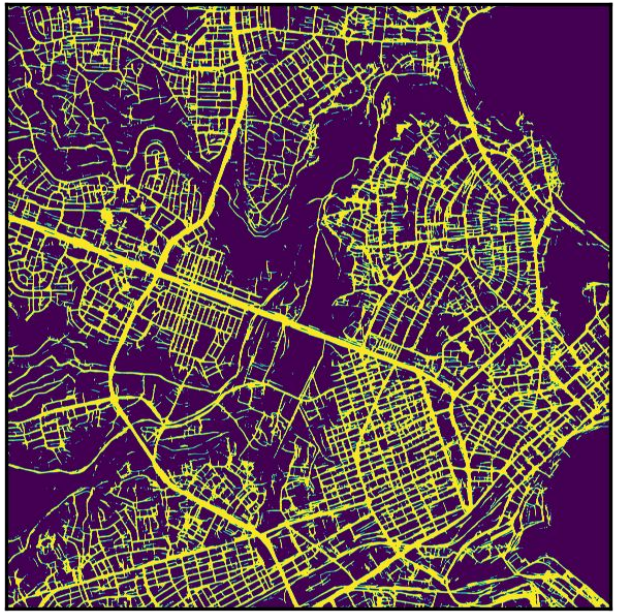
Play with ultra-large image

Large satellite images (up to ~40k×50k pixels) are sliced into 1300x1300 chips for mask prediction. After prediction, we stitch the masks back to a large image

Extract road network from predict mask

RNG image of 13000^2 pixels

Predicted mask of 13000^2 pixels



Extracted road network

Predicted mask can used to extract the road network graph.

Speed inference and output file write

	ImageId	WKT_Pix	length_m	travel_time_s
0	SN5_roads_test_public_AOI_7_Moscow_chip101	LINESTRING (32 130, 33 136, 42 176, 66 254, 70...	343.038923	27.517
1	SN5_roads_test_public_AOI_7_Moscow_chip101	LINESTRING (293 1109, 318 1212, 332 1259, 334 ...	54.622219	5.405
2	SN5_roads_test_public_AOI_7_Moscow_chip177	LINESTRING (589 0, 574 7, 515 105, 488 143, 46...	66.813805	9.184
3	SN5_roads_test_public_AOI_7_Moscow_chip177	LINESTRING (461 177, 413 176, 373 178, 343 183...	32.065023	4.174
4	SN5_roads_test_public_AOI_7_Moscow_chip177	LINESTRING (461 177, 468 182, 591 217, 633 227...	72.694431	9.695

Implications and Future

- Dataset with better annotation
- Include other released SpaceNet dataset in training could improve our models
- Tune the losses function weight to enhance predictions for speed channels
- Try various encoder structures for Unet backbone

References

[1]: SpaceNet 5 Challenge