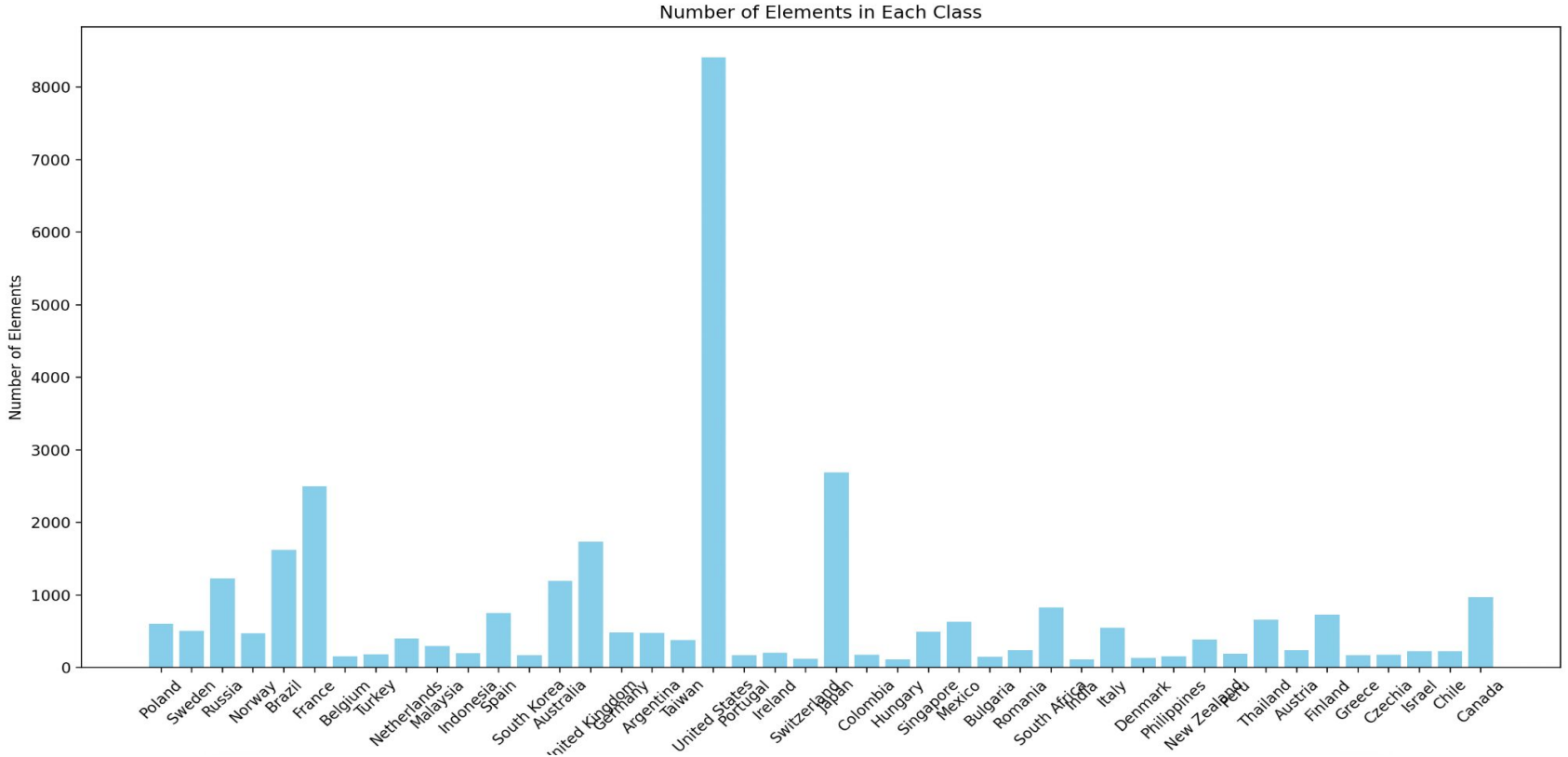
Where am I? AI for Geolocation

Maximilien Bohm,
Mathieu Gierski

Task 1 - Worldwide image classification by country

1. Data Imbalance



Initial number of images in class

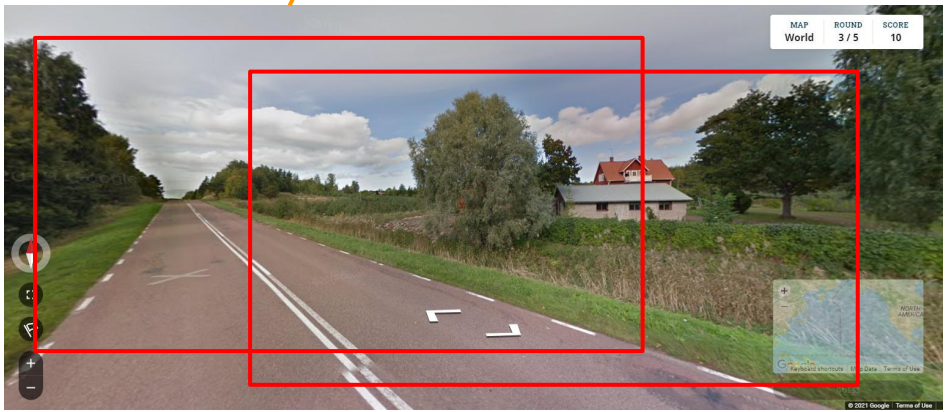
Symmetric images generated

Images with black holes added

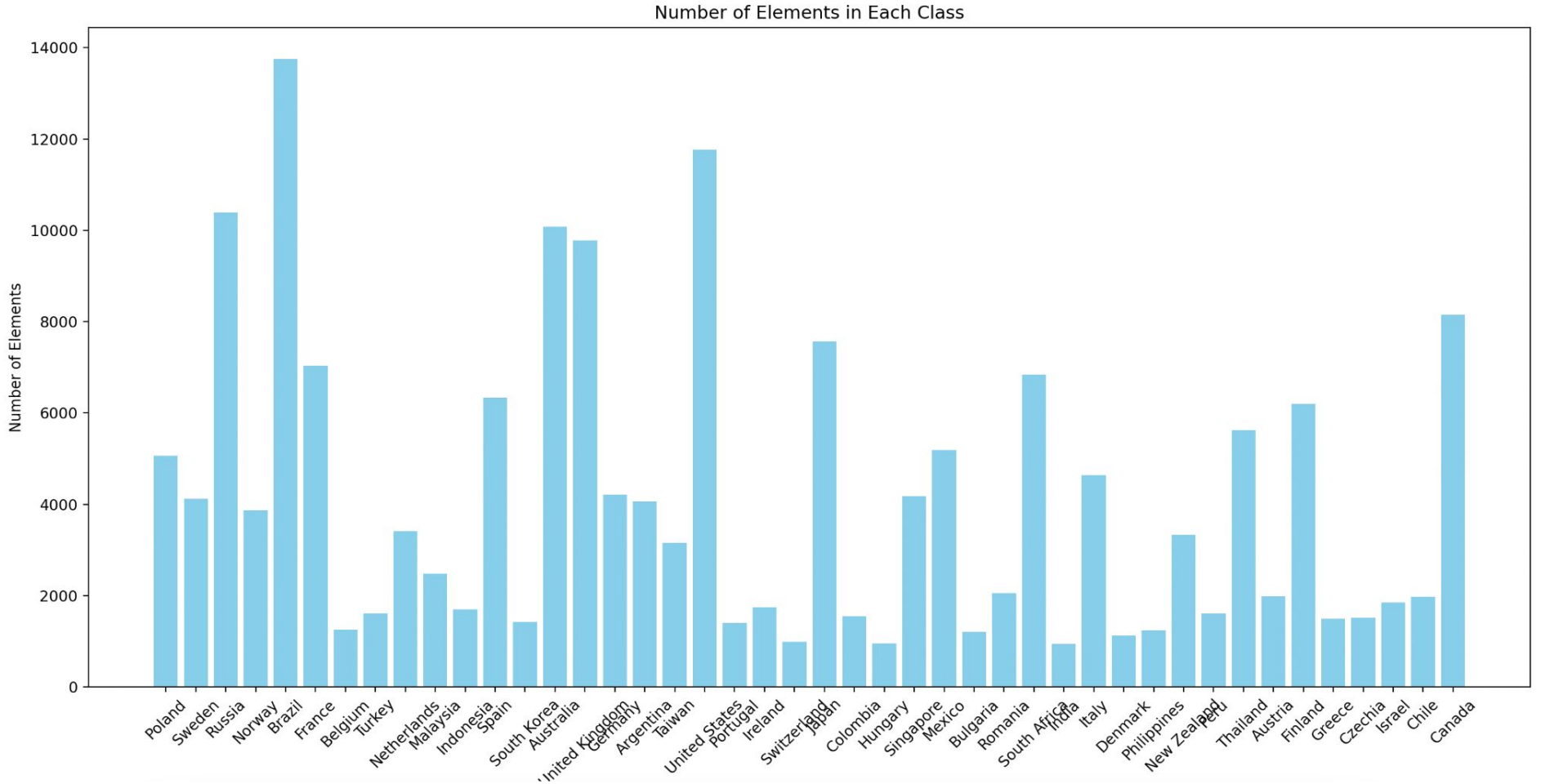
$$x(1+a)(1+b)(1+c+d)$$

Resampling generated

Gaussian noise added



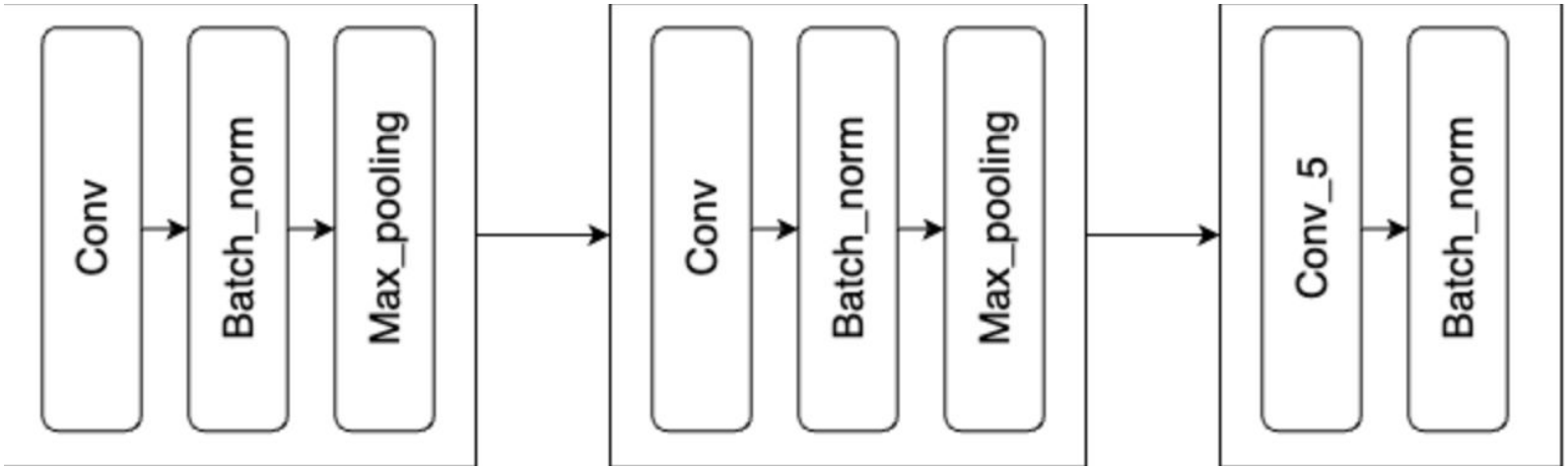
Mean new amount of images per class computed from initial amount x in each class.


Number of Elements in Each Class
Number of Elements
0
2000
4000
6000
8000
10000
12000
14000
Poland
Sweden
Russia
Norway
Brazil
France
Belgium
Turkey
Netherlands
Malaysia
Indonesia
Spain
South Korea
Australia
United Kingdom
Germany
Argentina
Taiwan
United States
Portugal
Ireland
Switzerland
Japan
Colombia
Hungary
Singapore
Mexico
Bulgaria
Romania
South Africa
India
Italy
Denmark
Philippines
New Zealand
Peru
Thailand
Austria
Finland
Greece
Czechia
Israel
Chile
Canada

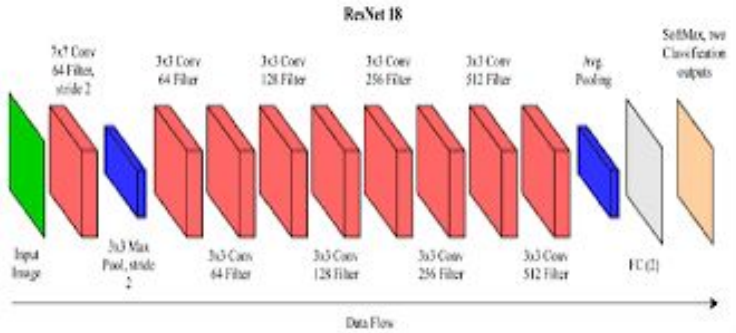
2. Models used

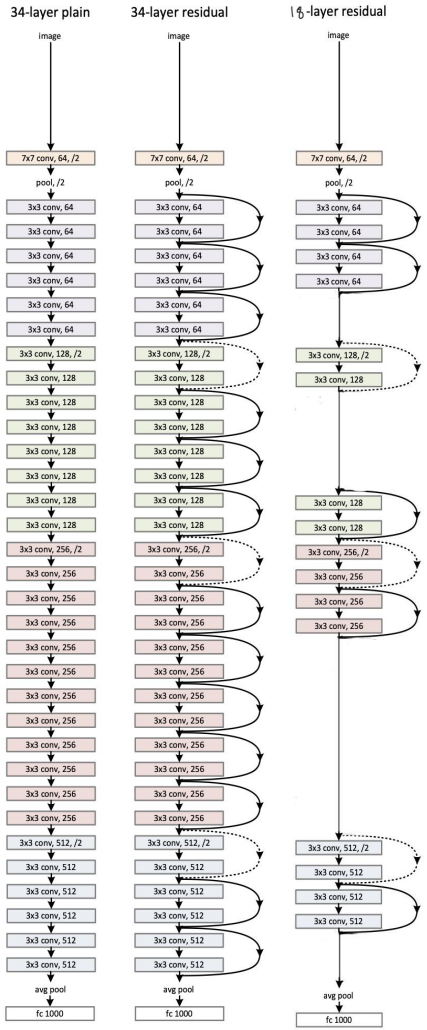
- CNN Model trained from scratch
- ResNet-18 Architecture trained from scratch
- Vision Transformer Architecture trained from scratch
- ~~Fine-tuning of Vision Transformer Model~~

CNN Architecture

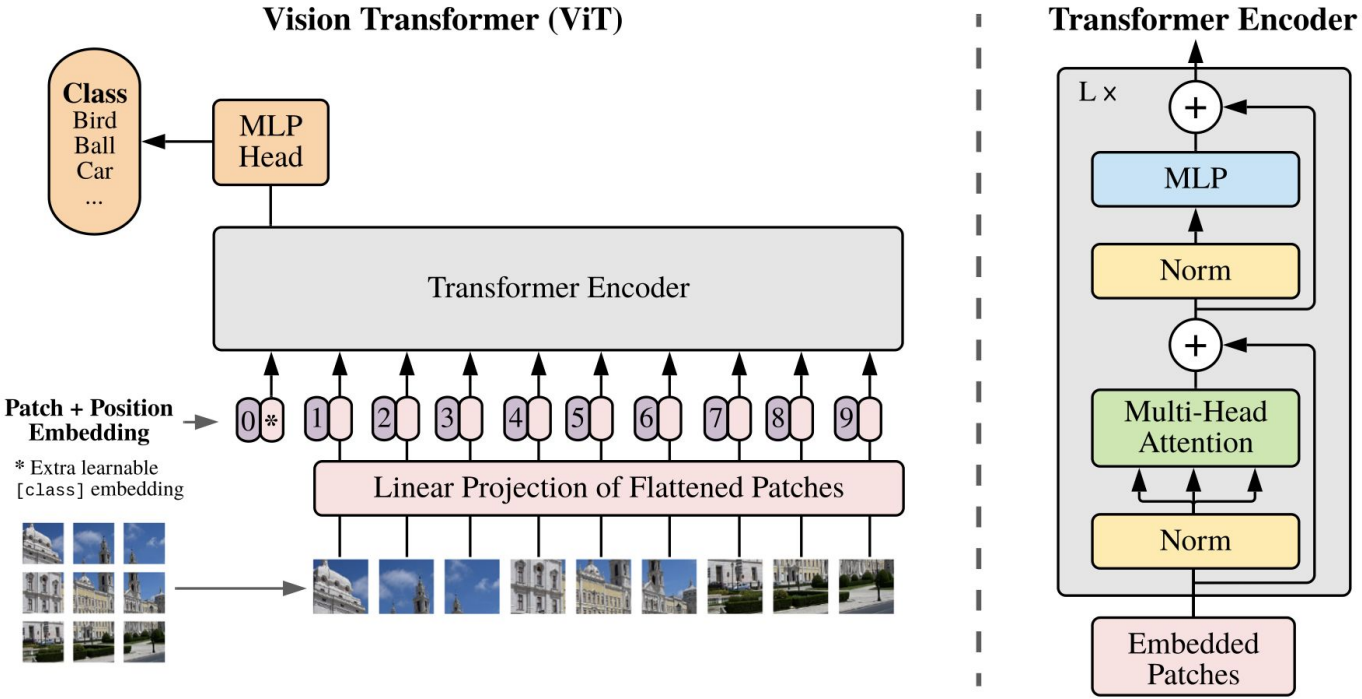


ResNet-18 Architecture

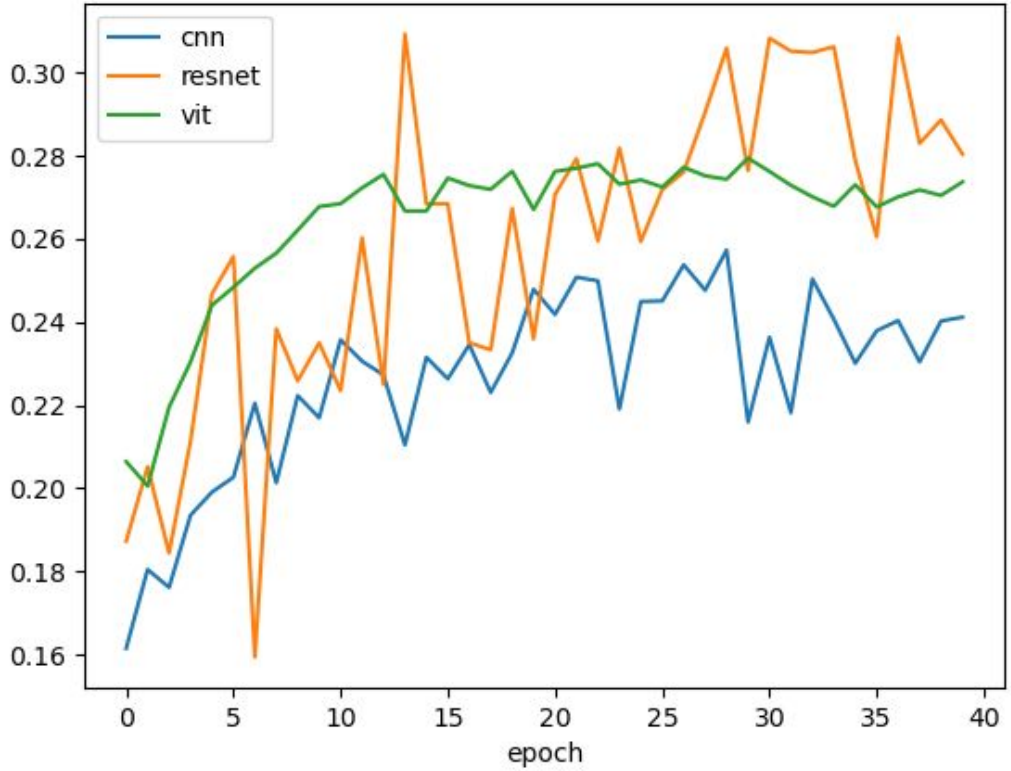




ViT Architecture



Accuracy (CNN, ResNet, ViT)



Pretrained ViT

`google/vit-base-patch16-224`

- Far too long to train
- Far too big model

Task 2 - Precise location regression within a city

Only 400 images





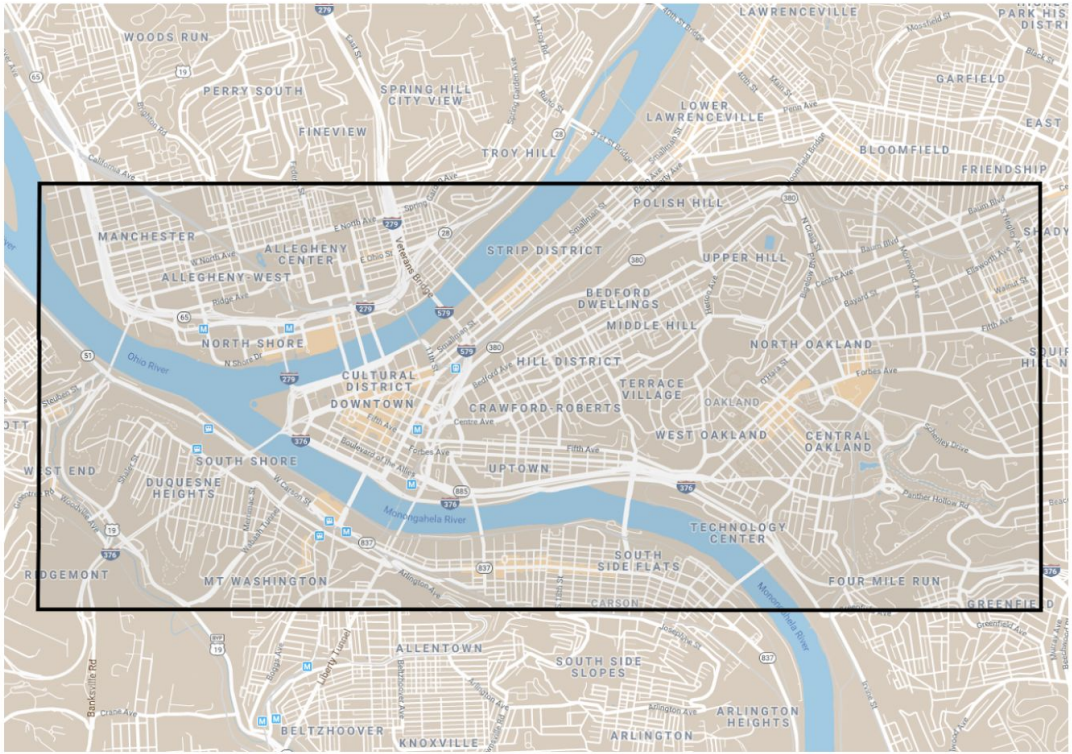

Correspondance Matrix

StreetLearn dataset (Google DeepMind):

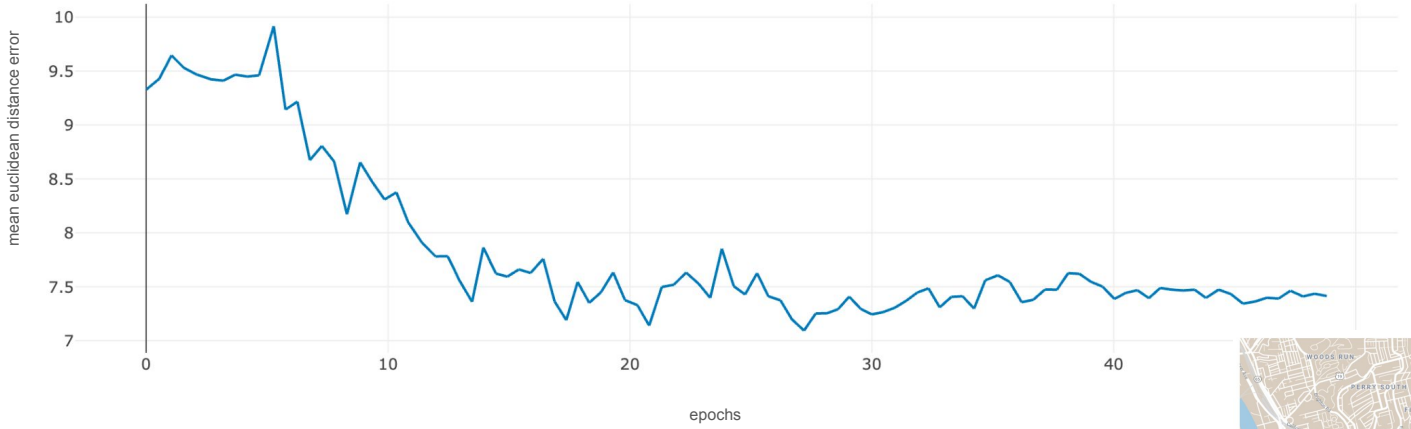
- 58 000 images
- within bounding box: (40.425, -80.035), (40.460, -79.930)
- Covers an area of 8.9km×3.9km or 34.3km^2

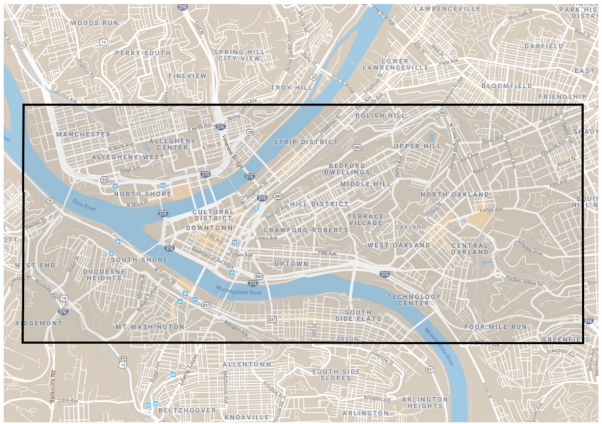
longitude	latitude	angle
40.440308999999999	-80	115.73999999999999
40.440271000000003	-80.006799999999998	119.23999999999999
40.440229000000002	-80.006699999999995	118.95999999999999
40.440188999999997	-80.006600000000006	118.68000000000001
40.440145999999999	-80.006500000000003	118.40000000000001
40.440094999999999	-80	118.12
40.440054000000003	-80.006200000000007	117.84
40.440009000000003	-80.006	117.56
40.439962000000001	-80.006	117.27
40.439914000000002	-80.005799999999994	116.98999999999999
40.439866000000002	-80.005700000000004	116.70999999999999
40.439818000000002	-80.005600000000001	116.43000000000001
40.439768999999998	-80.005399999999995	116.15000000000001
40.439720000000001	-80.005300000000005	115.87
40.439672000000002	-80.005200000000002	115.59
40.439605	-80.004999999999995	115.31
40.439523999999999	-80.004800000000003	115.02
40.439475000000002	-80.004000000000005	114.73999999999999
40.439427999999999	-80.004499999999993	114.45999999999999
40.439380999999997	-80.004400000000004	114.18000000000001
40.439334000000002	-80.004300000000001	113.90000000000001
40.43929	-80.004199999999997	113.62
40.439245999999997	-80.004000000000005	113.34
40.4392	-80.003900000000002	113.06
40.439157000000002	-80.003799999999998	112.77
40.439115000000001	-80.003699999999995	112.48999999999999
40.439064000000002	-80.003600000000006	112.20999999999999
40.439008000000001	-80.003399999999999	111.93000000000001
40.438958999999997	-80.003299999999996	111.65000000000001
40.438907999999998	-80.003100000000003	111.37
40.438859000000001	-80.003	111.09
40.438814999999998	-80.002899999999997	110.81
40.438775999999997	-80.002799999999993	110.52
40.438732000000002	-80.002600000000001	110.23999999999999
40.438692000000003	-80.002499999999998	109.95999999999999

Images covering areas of Pittsburgh



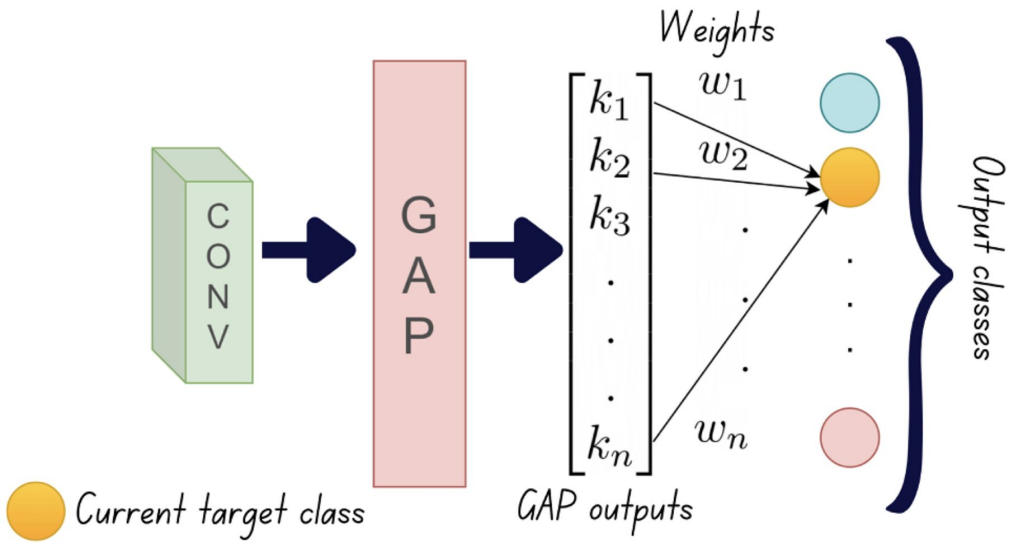
Evaluating on our models





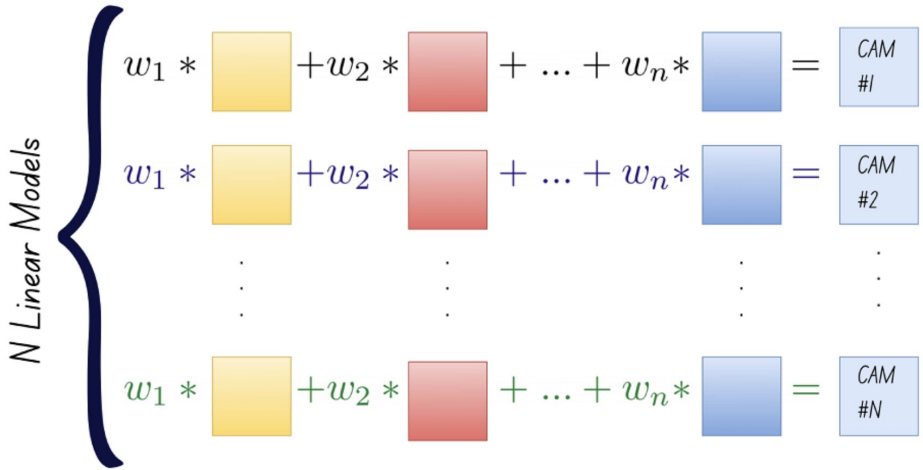
- **Inputs:** Image inputs (240x120)
- **Outputs:** 2 numbers (Long, Lat) (normalized later denormalized)

Task 3 - Class Activation Maps



Tried on a CNN:

- after **Conv_2** layer (image of size **(5*10)**, depth **40**)
- after **Conv_3** layer (images of size **(2*5)**, depth **60**)



In Poland

	After **conv_2**	After **conv_3**
CAM for ***Poland***	 *0%*	 *0%*
CAM of ***Predicted country***	 *New Zealand 23%*	 *New Zealand 66%*