

Landmark Classification

EGH444 - Digital Signals & Image Processing



Lachlan Kuhr, n9767151

Andrew Mather, n9713671

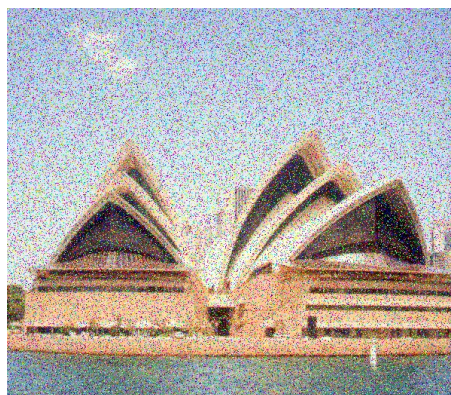
Jeremy Sheather, n9734783

October, 2020

Data Collection & Augmentation

- Images selected from Google Landmark Database
 - Wide range of angles, perspectives, weather conditions etc
- 300 base 'Opera House' images, 400 base 'Eiffel Tower' images, 400 randomly selected base 'Other' images, all with various augmentations

Eiffel Tower - Rotated



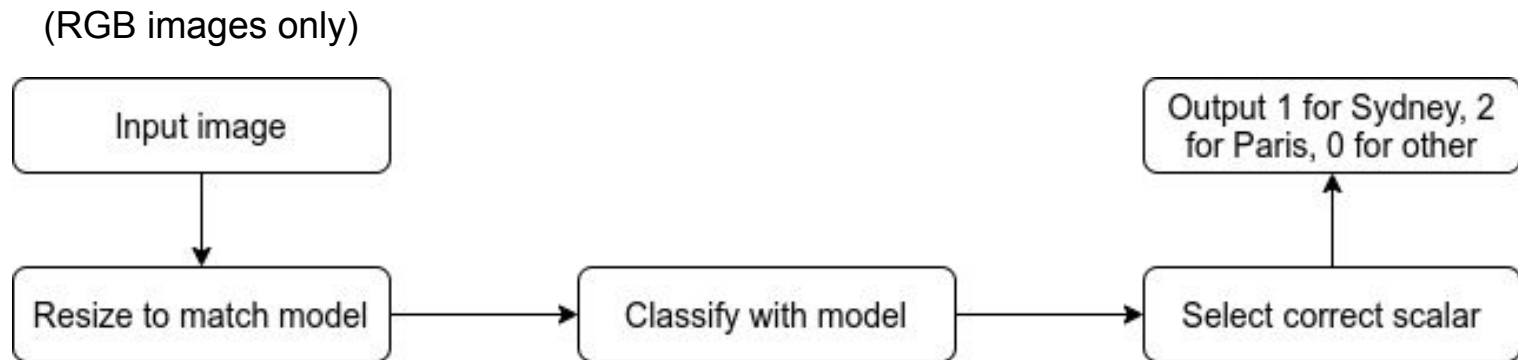
Opera House - Salt and Pepper Noise

Eiffel Tower - Gaussian Noise, Rotated



Opera House - Geometric Distortion

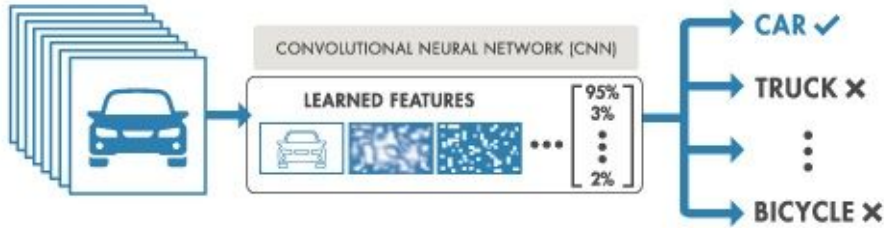
Classification Method



Classification Pipeline

Model Research & Experimentation

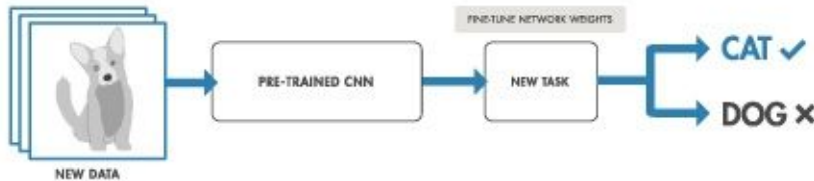
TRAINING MODEL FROM SCRATCH



From Scratch Procedure:

1. Define layers
2. Train the network on the data
3. Test performance

TRANSFER LEARNING



Transfer Learning Procedure:

1. Load a pretrained network
2. (Optional) Add additional layers
3. Replace the classification layers for the task
4. Train the network on the data
5. Test performance

Transfer Learning Considerations

- Pretrained models
 - AlexNet, ResNet, InceptionV3, MobileNetV2, VGG-16
 - Tradeoff between network accuracy, speed, and size
- Pretrained model dataset
 - ImageNet or Places365
- Adding extra layers
- Freezing weights
- Training/validation dataset split
- Hyperparameters tuning
 - Learning rate, batch size, epochs

Selection Of Models

Pretrained Model	Frozen Layers	Model Size Memory Parameters	Validation Dataset	Test Set 1 (standard)	Test Set 2 (harder)
AlexNet	0	200MB 61M	95.0%	97.8%	84.2%
ResNet-18	4	40MB 11M	96.3%	97.8%	92.5%
MobileNetV2	16	8MB 3.4M	95.9%	100%	92.5%
InceptionV3	17	80MB 24M	96.8%	97.8%	93.2%

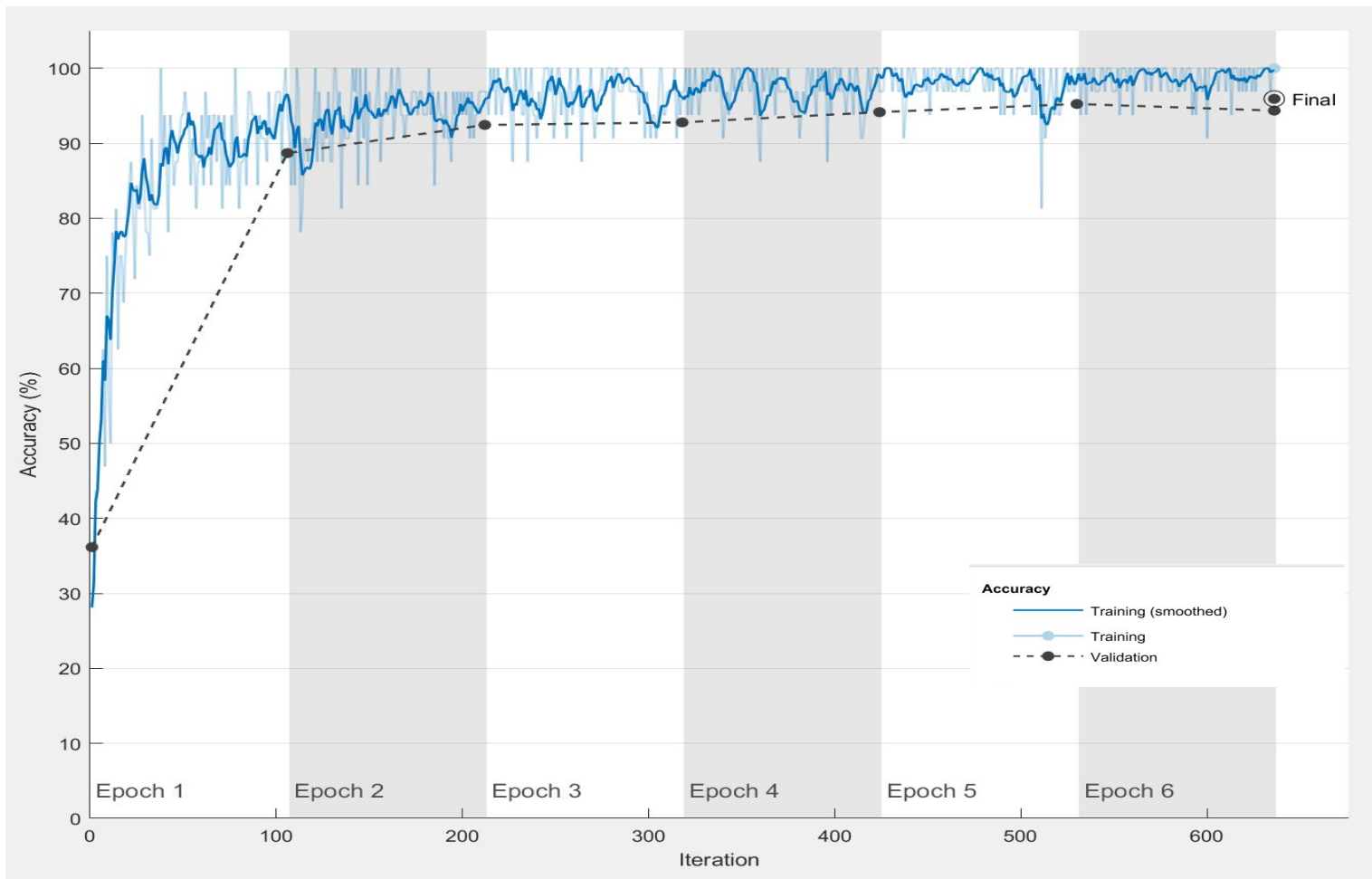
Test datasets:

- Validation (1466) → 30% reserved from original training dataset
- Test set 1 (90) → Standard images
- Test set 2 (150) → Harder images

Final Model

- Pretrained model: InceptionV3
 - Balance of accuracy and model size
- Final two layers replaced (FCL & output)
- Hyper-parameters:
 - 6 epochs
 - Batch size = 32
 - Learning rate low but high for last two layers

Training & Validation Accuracy



Model Performance

- Dataset accuracy
 - Validation accuracy: 96.8%
 - Standard images accuracy: 97.8%
 - Hard images accuracy: 93.2%
- Incorrect Classifications:
 - Most errors occur on level 2 and 3 images
 - Level 1 errors: Landmark small and in background

Confusion Matrix

True Class	Predicted Class			Percentage	
	Other	Paris	Sydney		
Other	512	5	5	98.1%	1.9%
Paris	15	536	1	97.1%	2.9%
Sydney	18	3	371	94.6%	5.4%

- Most errors occur as “Paris” and “Sydney” classified as “Other”
 - 15 “Paris” images
 - 18 “Sydney” images
- These images tend to only feature the landmark in the background

Challenging, Correct Classification

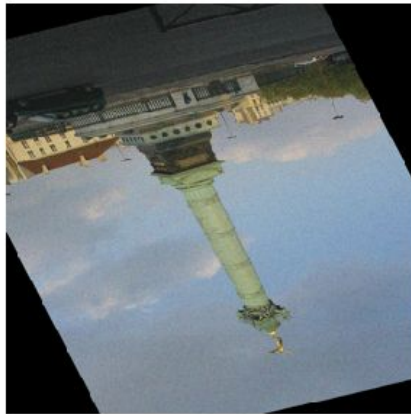
Paris



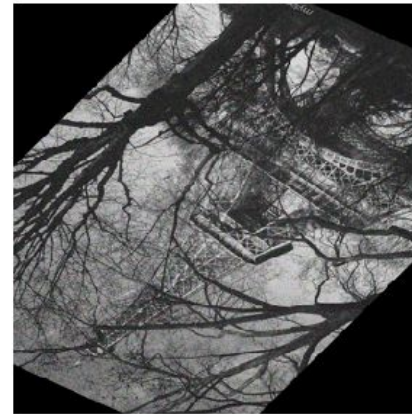
Sydney



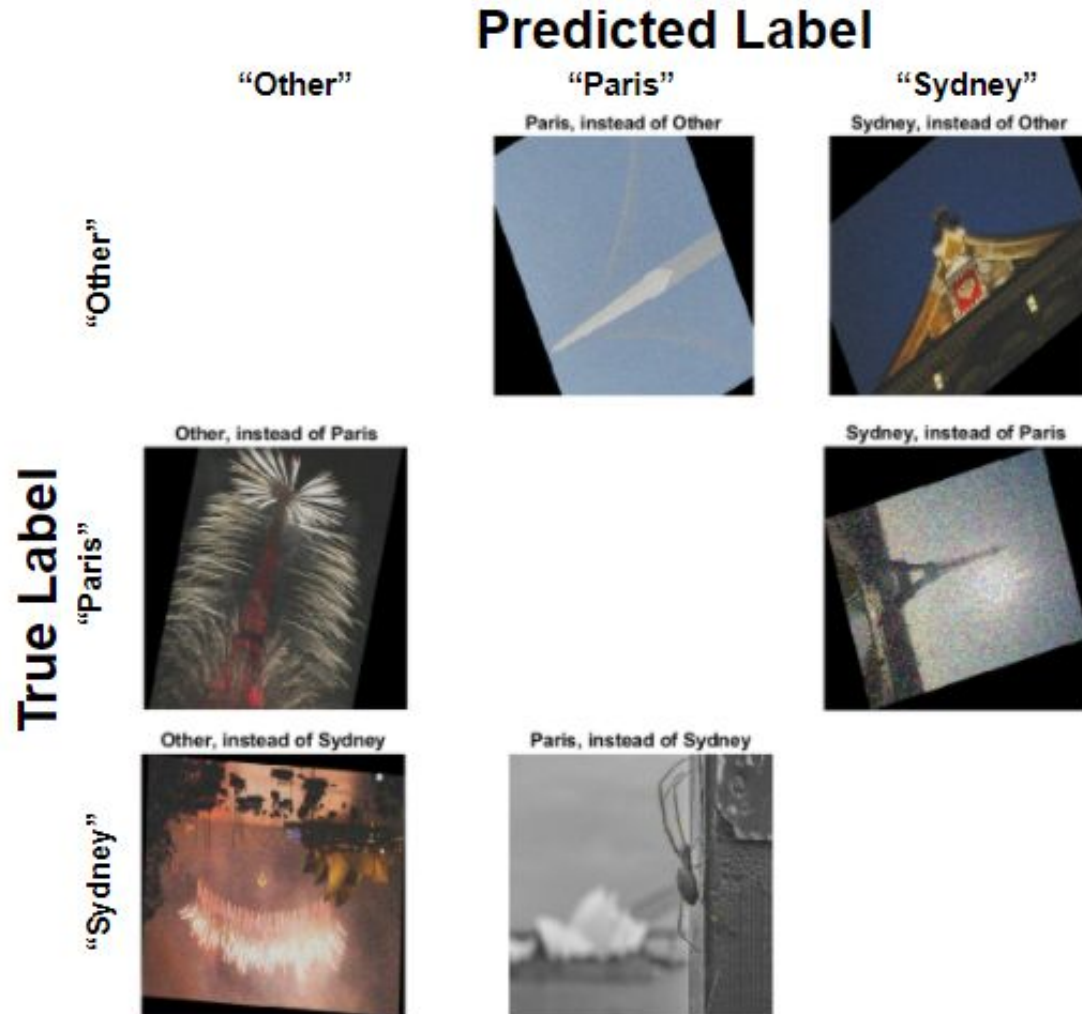
Other



Paris



Incorrect Classifications



Model Limitations & Recommendations

Limitations:

- Landmark in background tends to perform poorly
- Struggles on rotated and cropped images
- Relatively small training dataset

Recommendation:

- Bayesian Deep Learning to quantify uncertainty
- Target detection and bounding boxes using R-CNN

Questions