Visual Concept Detection Machine learning challenge

Nooshin Shojaee , Lucas Souza Romao , Francesco Ciraolo

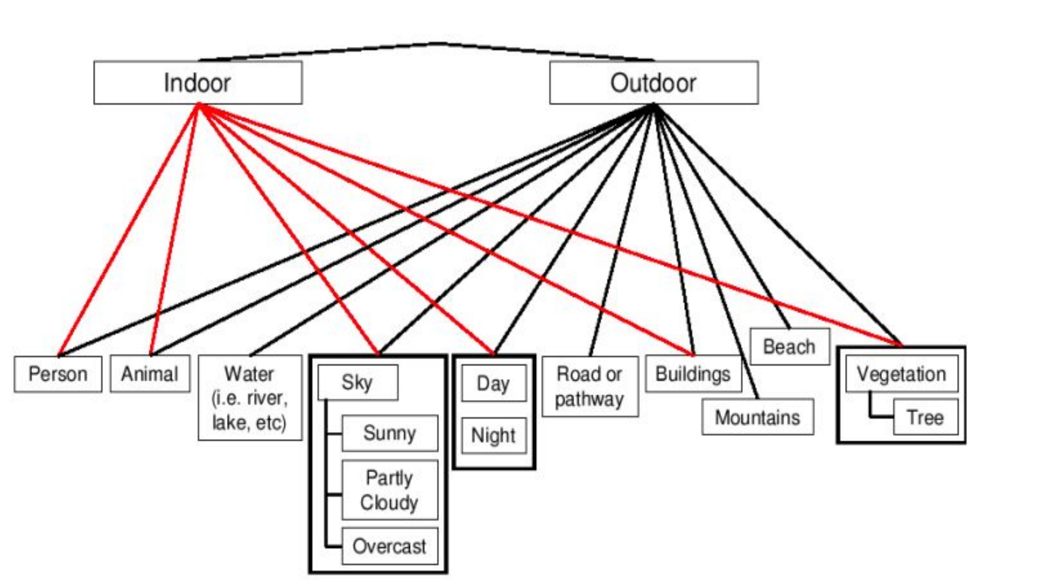
Professor : Dr. Vladimir Despotovic

Out lines

- The problem
- The data set
- Libraries
- Data preparation
- Model
- Results

Visual Concept Detection

Identify the presence/absence of 17 visual concepts in Images



- **Format of the annotation**

27-27700.jpg 0 1 0 0 1 0 0 1 1 1 0 0 1 0 0 0 0
27-27704.jpg 0 1 0 1 0 0 0 0 0 1 0 0 1 0 1 0 0
27-27705.jpg 0 1 0 0 1 0 0 1 0 1 0 0 1 0 0 1 0
27-27706.jpg 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
27-27709.jpg 0 1 0 1 0 0 1 0 0 0 0 0 1 0 0 1 0
27-27712.jpg 0 1 0 0 0 0 0 1 0 1 0 0 1 0 1 0 0

IAPR TC-12 Data set

Tensorflow+Keras

Tensorflow :

Google Open-source library for developing and training machine learning models

Keras :

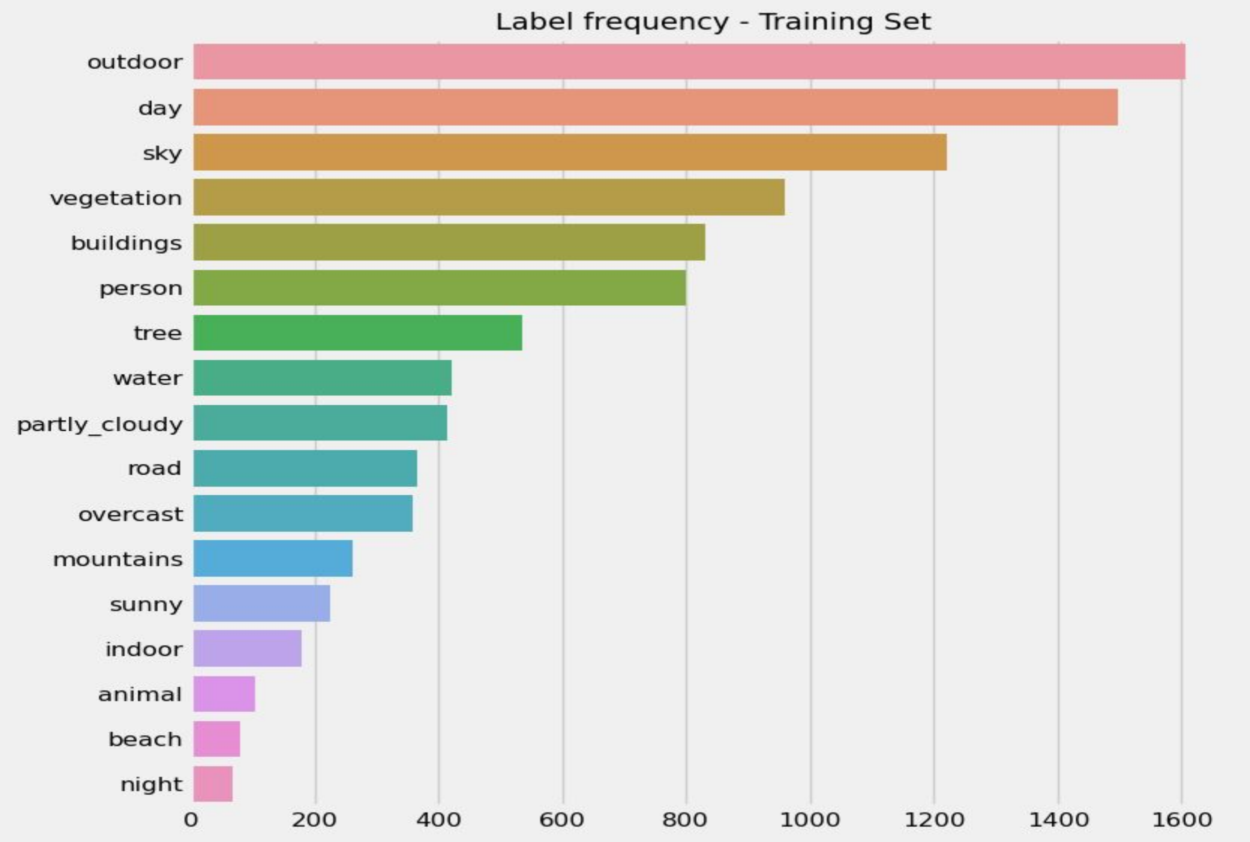
Acts as an interface for the TensorFlow library

Data preparation Challenges

1. **GIF types**

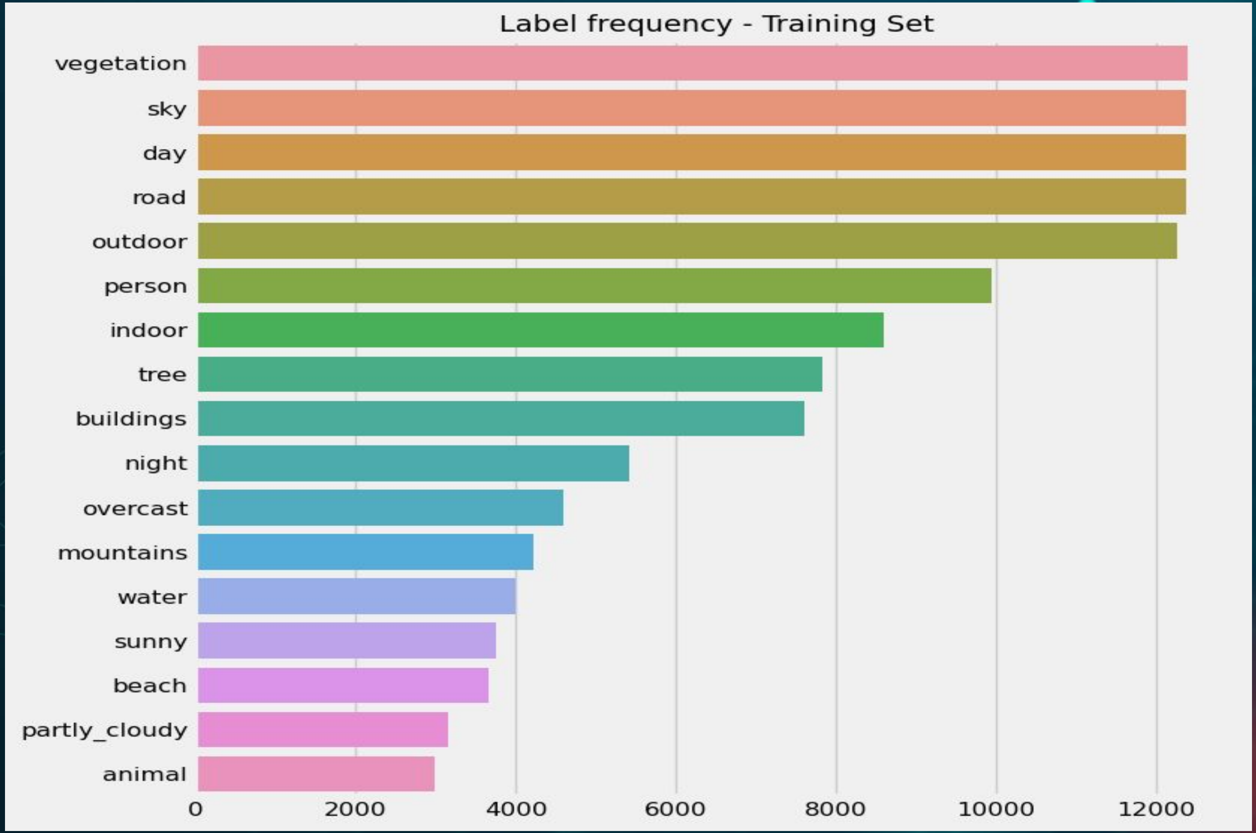
   Removed from the train dataset because they were less than 1%. (3 images)

2. **Unbalanced data set**



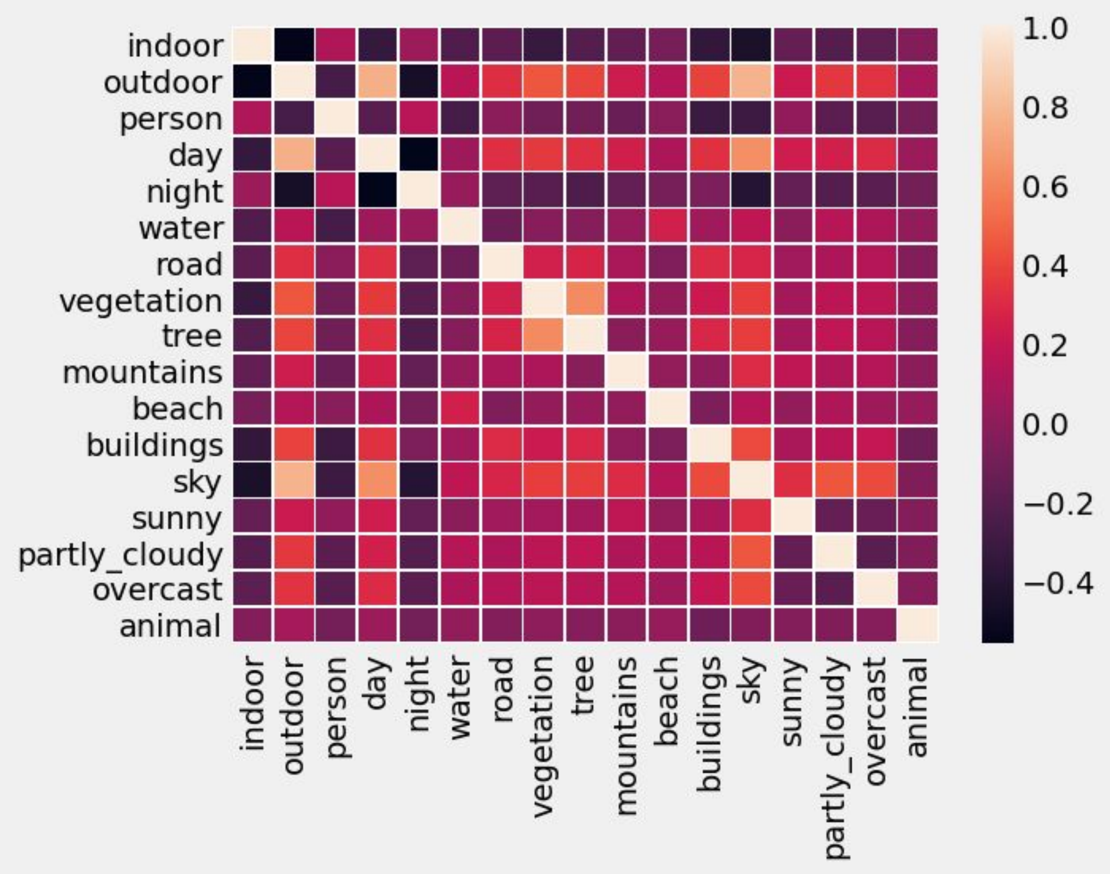
Data preparation Challenges

- SMOTETomek



Data preparation Challenges

Features are correlated so oversampling and undersampling weren't effective in the results.



Data preparation Challenges

Classes Weights

'indoor': 2.12
'Outdoor': 1.0
'person': 1.0
'day': 1.0
'night': 3.0
'water': 1.26,
'road': 1.40
'vegetation': 1.0
'tree': 1.02
'mountains': 1.74
'beach': 2.92

'buildings': 1.0
'sky': 1.0
'sunny': 1.89
'partly_cloudy': 1.28
'overcast': 1.42
'animal': 2.67

Data Argumentation

Vertical flip

Shear

Rotation

Zoom

Horizontal flip

Model choice

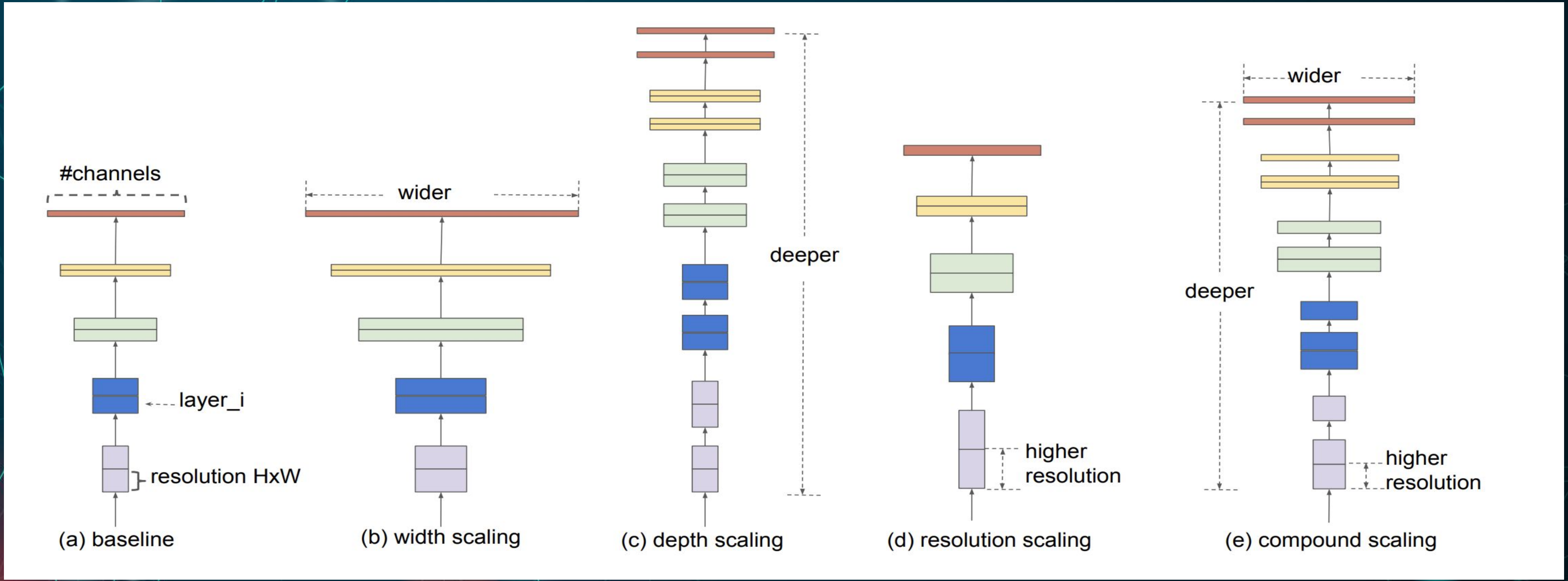
Tested models

- ResNet 50
- EfficientNet B1
- EfficientNet B2

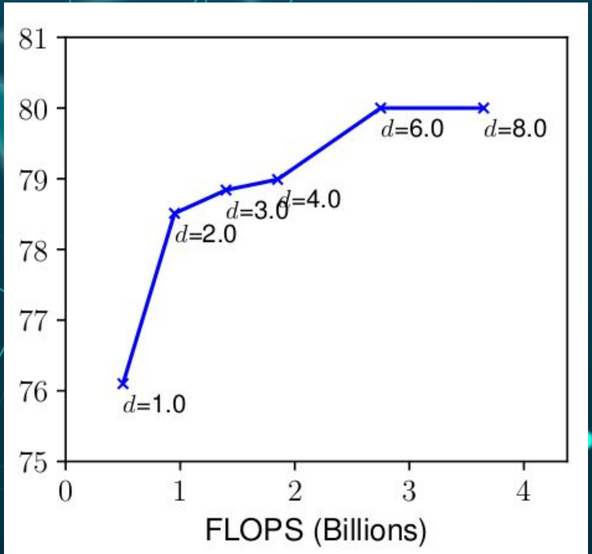
Selected

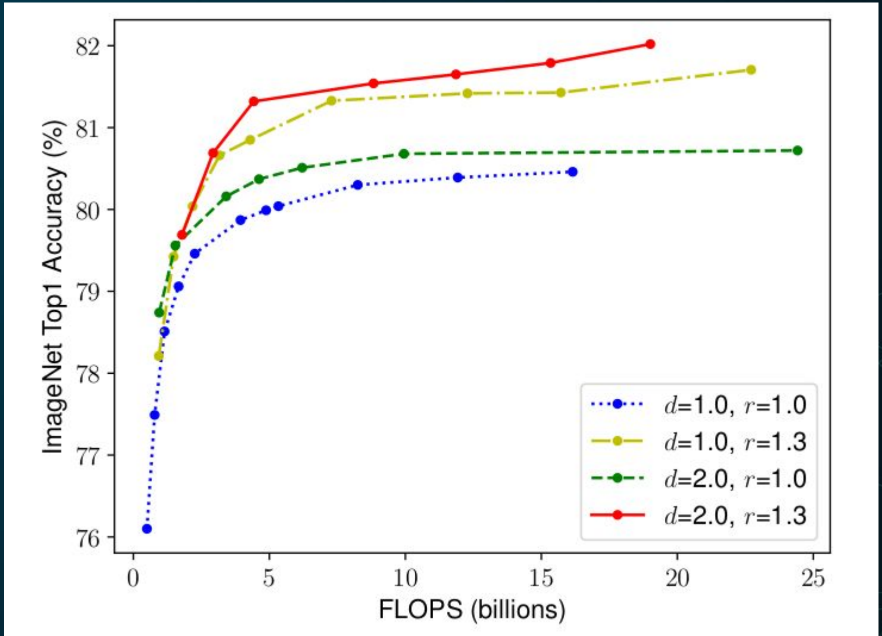
- EfficientNet B2

ResNet vs EfficientNet - I



ResNet vs EfficientNet - II





ResNet vs EfficientNet - III

Model	Top-1 Acc.	Top-5 Acc.	#Params	Ratio-to-EfficientNet	#FLOPs	Ratio-to-EfficientNet
EfficientNet-B0	**77.1%**	**93.3%**	**5.3M**	**1x**	**0.39B**	**1x**
ResNet-50 (He et al., 2016)	76.0%	93.0%	26M	4.9x	4.1B	11x
DenseNet-169 (Huang et al., 2017)	76.2%	93.2%	14M	2.6x	3.5B	8.9x
EfficientNet-B1	**79.1%**	**94.4%**	**7.8M**	**1x**	**0.70B**	**1x**
ResNet-152 (He et al., 2016)	77.8%	93.8%	60M	7.6x	11B	16x
DenseNet-264 (Huang et al., 2017)	77.9%	93.9%	34M	4.3x	6.0B	8.6x
Inception-v3 (Szegedy et al., 2016)	78.8%	94.4%	24M	3.0x	5.7B	8.1x
Xception (Chollet, 2017)	79.0%	94.5%	23M	3.0x	8.4B	12x
EfficientNet-B2	**80.1%**	**94.9%**	**9.2M**	**1x**	**1.0B**	**1x**
Inception-v4 (Szegedy et al., 2017)	80.0%	95.0%	48M	5.2x	13B	13x
Inception-resnet-v2 (Szegedy et al., 2017)	80.1%	95.1%	56M	6.1x	13B	13x
EfficientNet-B3	**81.6%**	**95.7%**	**12M**	**1x**	**1.8B**	**1x**
ResNeXt-101 (Xie et al., 2017)	80.9%	95.6%	84M	7.0x	32B	18x
PolyNet (Zhang et al., 2017)	81.3%	95.8%	92M	7.7x	35B	19x
EfficientNet-B4	**82.9%**	**96.4%**	**19M**	**1x**	**4.2B**	**1x**
SENet (Hu et al., 2018)	82.7%	96.2%	146M	7.7x	42B	10x
NASNet-A (Zoph et al., 2018)	82.7%	96.2%	89M	4.7x	24B	5.7x
AmoebaNet-A (Real et al., 2019)	82.8%	96.1%	87M	4.6x	23B	5.5x
PNASNet (Liu et al., 2018)	82.9%	96.2%	86M	4.5x	23B	6.0x
EfficientNet-B5	**83.6%**	**96.7%**	**30M**	**1x**	**9.9B**	**1x**
AmoebaNet-C (Cubuk et al., 2019)	83.5%	96.5%	155M	5.2x	41B	4.1x

EfficientNet B0 architecture

Stage i	Operator $\hat{F}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224 × 224	32	1
2	MBConv1, k3x3	112 × 112	16	1
3	MBConv6, k3x3	112 × 112	24	2
4	MBConv6, k5x5	56 × 56	40	2
5	MBConv6, k3x3	28 x 28	80	3
6	MBConv6, k5x5	14 x 14	112	3
7	MBConv6, k5x5	14 x 14	192	4
8	MBConv6, k3x3	7 x 7	320	1
9	Conv1x1 & Pooling & FC	7 x 7	1280	1

Class	precision	recall	f1 score
indoor	0.66	0.68	0.67
outdoor	0.95	0.95	0.95
person	0.63	0.65	0.64
day	0.88	0.95	0.92
night	0.00	0.00	0.00
water	0.47	0.16	0.24
road	0.00	0.00	0.00
vegetation	0.74	0.77	0.76
tree	0.66	0.55	0.60
mountains	0.50	0.01	0.01
beach	0.00	0.00	0.00
buildings	0.67	0.67	0.67
sky	0.94	0.84	0.88
sunny	0.61	0.39	0.47
partly_cloudy	0.52	0.24	0.33
overcast	0.76	0.45	0.57
animal	0.00	0.00	0.00

Average precision: 0.80
Average recall: 0.67
Average f1 score: 0.73

ResNet50

Class	precision	recall	f1 score
indoor	0.68	0.69	0.68
outdoor	0.96	0.96	0.96
person	0.64	0.54	0.59
day	0.90	0.96	0.93
night	0.00	0.00	0.00
water	0.50	0.29	0.36
road	0.70	0.04	0.07
vegetation	0.73	0.83	0.78
tree	0.56	0.55	0.56
mountains	0.48	0.07	0.13
beach	0.00	0.00	0.00
buildings	0.67	0.69	0.68
sky	0.93	0.93	0.93
sunny	0.58	0.60	0.59
partly_cloudy	0.56	0.39	0.46
overcast	0.76	0.34	0.47
animal	1.00	0.02	0.03

Average precision: 0.79
Average recall: 0.70
Average f1 score: 0.74

EfficientNet B1

Model Parameters

- Batch Size: 32

Efficient Net B2:

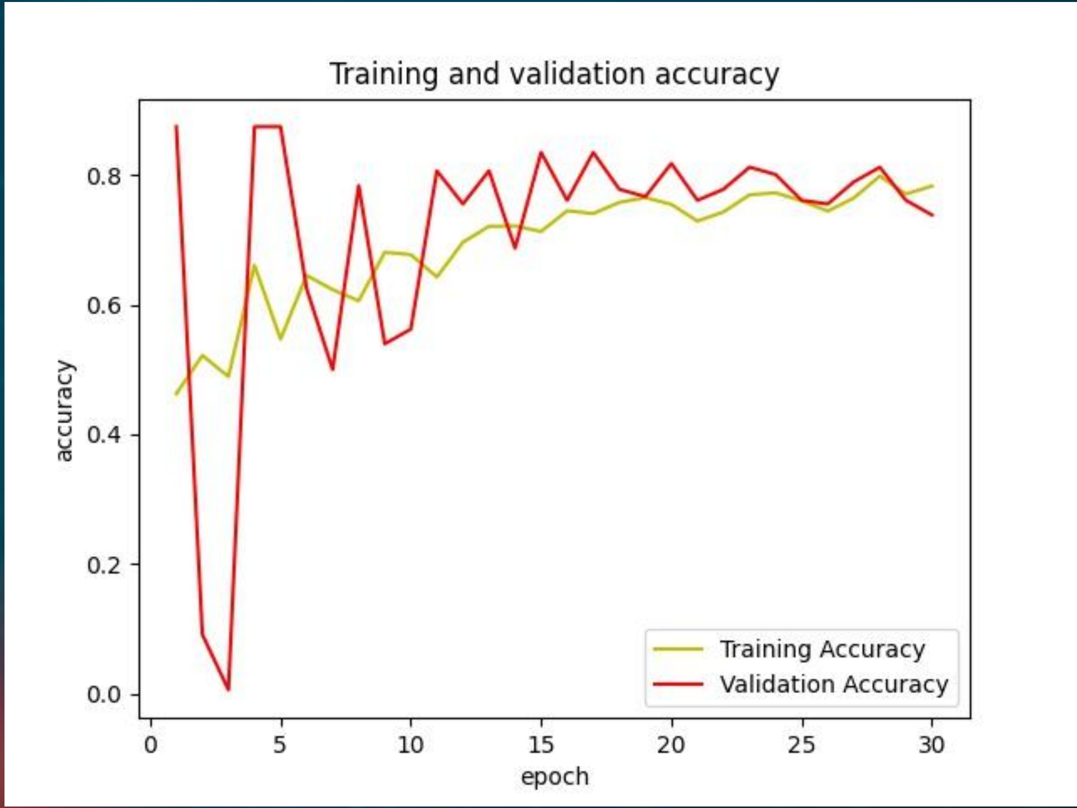
- Weights: None
- Pooling: Average
- DropOut Rate: 0.2
- Input Shape: 260 x 260 x 3

Model:

```
Layer (type)                 Output Shape              Param #
=================================================================
efficientnetb2 (Functional)  (None, 1408)              7768569

dense_1 (Dense)              (None, 512)               721408

dropout (Dropout)            (None, 512)               0

dense_2 (Dense)              (None, 17)                8721
=================================================================
Total params: 8,498,698
Trainable params: 8,431,123
Non-trainable params: 67,575
```

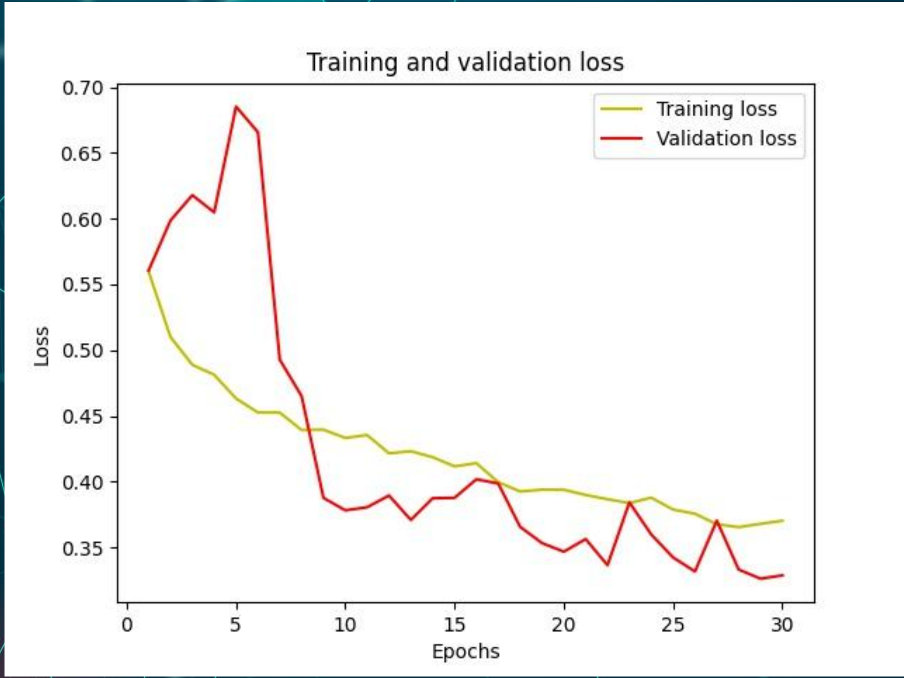
- Output Activation: Sigmoid
- Optimizer: Adam
- Loss: Binary Cross Entropy
- Callbacks: ReduceLROnPlateau

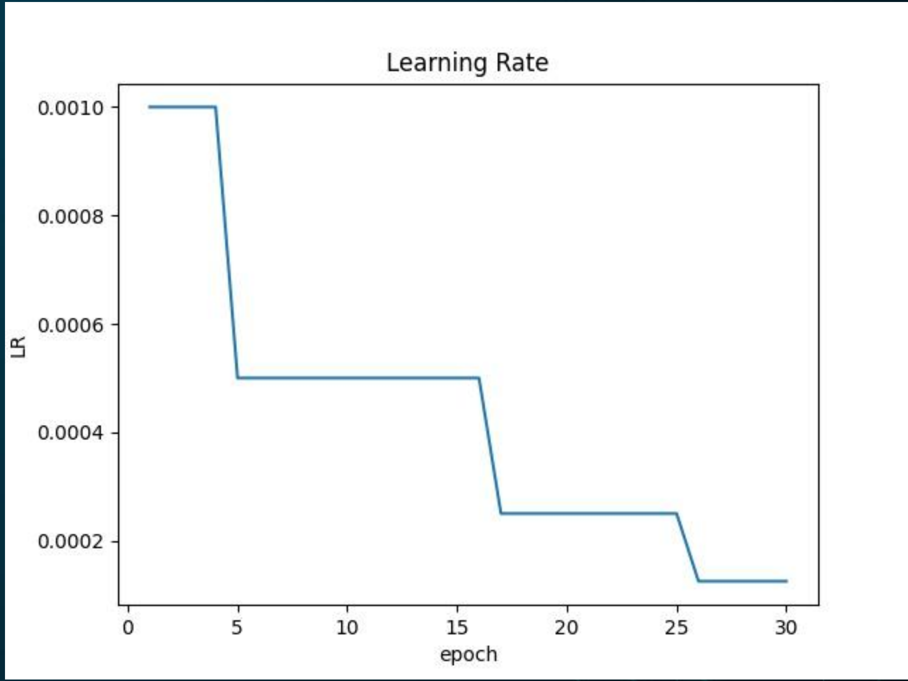
Results



Class	precision	recall	f1 score
indoor	0.88	0.35	0.50
outdoor	0.92	0.99	0.95
person	0.72	0.45	0.56
day	0.88	0.97	0.92
night	0.47	0.39	0.43
water	0.50	0.35	0.41
road	0.54	0.10	0.17
vegetation	0.72	0.86	0.78
tree	0.61	0.69	0.65
mountains	0.63	0.27	0.38
beach	0.00	0.00	0.00
buildings	0.70	0.76	0.73
sky	0.93	0.91	0.92
sunny	0.73	0.27	0.39
partly_cloudy	0.57	0.53	0.55
overcast	0.67	0.69	0.68
animal	0.60	0.05	0.10

Average precision: 0.79
Average recall: 0.73
Average f1 score: 0.76


Training and validation loss
Training loss
Validation loss
Loss
Epochs
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0
5
10
15
20
25
30


Learning Rate
LR
epoch
0.0010
0.0008
0.0006
0.0004
0.0002
0
5
10
15
20
25
30

Questions?