EPFL

Topic classification of Wikipedia images

Semester project at **DLAB**

Matheus Bernat

Supervised by
- Tiziano Piccardi
- Miriam Redi

July 2022

OUTLINE

Problems

Related work

Our work

Findings

The broader problems

1. Missing images
   - 10% of 6.5 million Wikipedia articles without images.[1]

Wikipedia requested images

2. Image vandalism
   - Between 2001 and 2014: 7% to 11% of 500 million edits were vandalism.[2]

The broader problems – proposed solution

Topic classification of Wikipedia images

https://commons.wikimedia.org/wiki/File:Leslie_J_Rissler_and_a_frog.jpg

Image ***content***

- *frog*
- *woman*
- *jacket*

Image ***topic***:

- Biology
- Science
- Biography

The broader problems – gains of proposed solution

Topic classification of Wikipedia images

1. Missing images
   ⇒ Better recommendation systems for editors
2. Image vandalism
   ⇒ Alert if the image's topic isn't related to the article's topic
3. Visual knowledge gaps
   ⇒ Fill up topics with missing images
4. Readers' interaction
   ⇒ Adapt image display depending on its topic

Challenges with topic classification

(besides having meaningful labels)

- Distinct image features within the same topic; e.g. **History** (Francesco's labels)

https://commons.wikimedia.org/wiki/File:Carnegie_Library_Solvay_NY.jpg

https://als.wikipedia.org/wiki/Chilperich_I.#/media/Datei:Chilperic_I_&_Fredegunde00.jpg

https://commons.wikimedia.org/wiki/File:Katharine_Rosse_in_%22Games%22_%281967%29.jpg

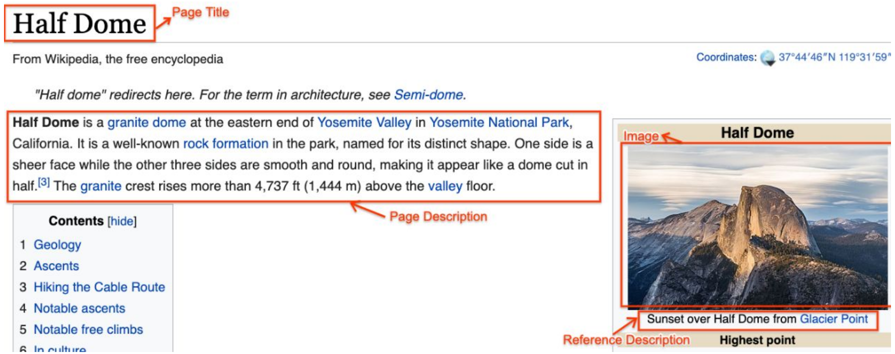
https://commons.wikimedia.org/wiki/File:Where_the_Twin_Towers_Were.jpg

OUTLINE

Related work

1. ORES is an ensemble of machine learning methods designed for Wikipedia, providing topic classification, text vandalism detection, etc.[5]

2. The Wikipedia-based Image Text (WIT) dataset contains 37.6M image-text entries, with 11.5M unique images.[4]



Related work

3. EfficientNet is a family of deep learning networks with better accuracy while requiring fewer parameters.[3]
4. Redi performed *image* classification of Wikipedia images linking 6.7M Commons categories to 160 COCO concepts. Fine-tuning network pre-trained on ImageNet.[6]

OUTLINE

Problems

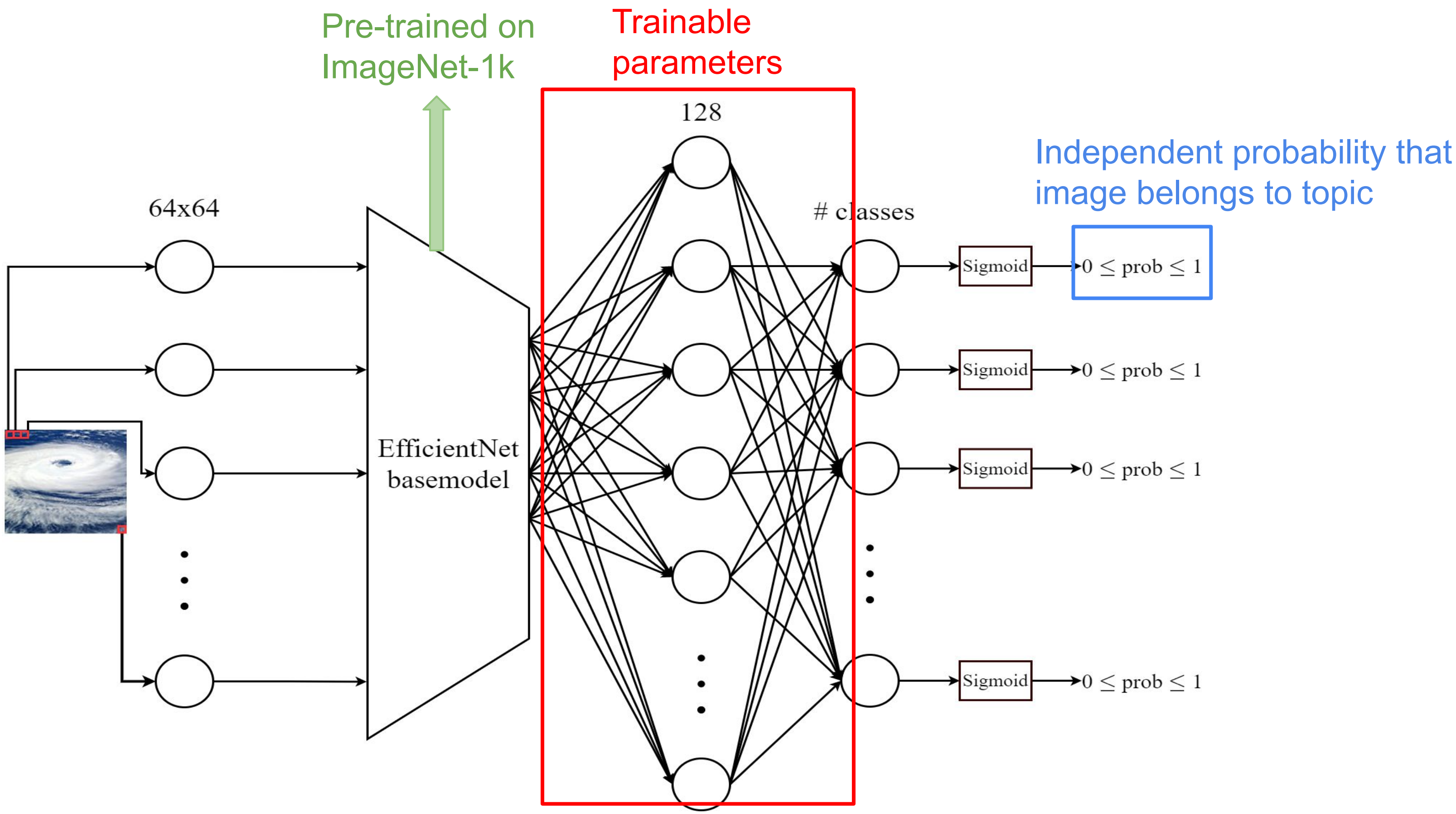
Related work

Our work

Findings

Our work

- Inspired in Redi's work, fine-tuning a deep learning network pre-trained on ImageNet
- Binary cross-entropy loss function for independent probabilities (i.e. an image can belong to more than 1 topic)
- Metrics: precision, recall, ROC AUC

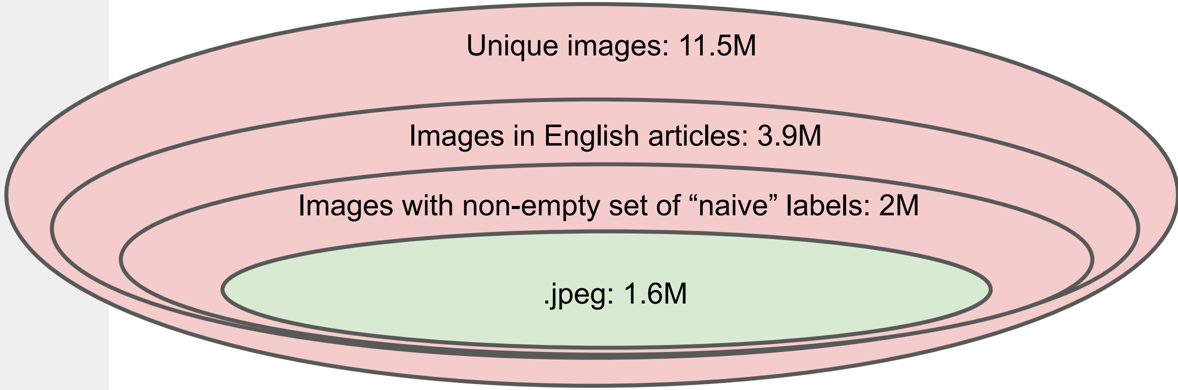

Pre-trained on ImageNet-1k
Trainable parameters
Independent probability that image belongs to topic
64x64
EfficientNet basemodel
128
classes
Sigmoid
$0 \leq \text{prob} \leq 1$
Sigmoid
$0 \leq \text{prob} \leq 1$
Sigmoid
$0 \leq \text{prob} \leq 1$
Sigmoid
$0 \leq \text{prob} \leq 1$

EPFL

Our work – data (images)

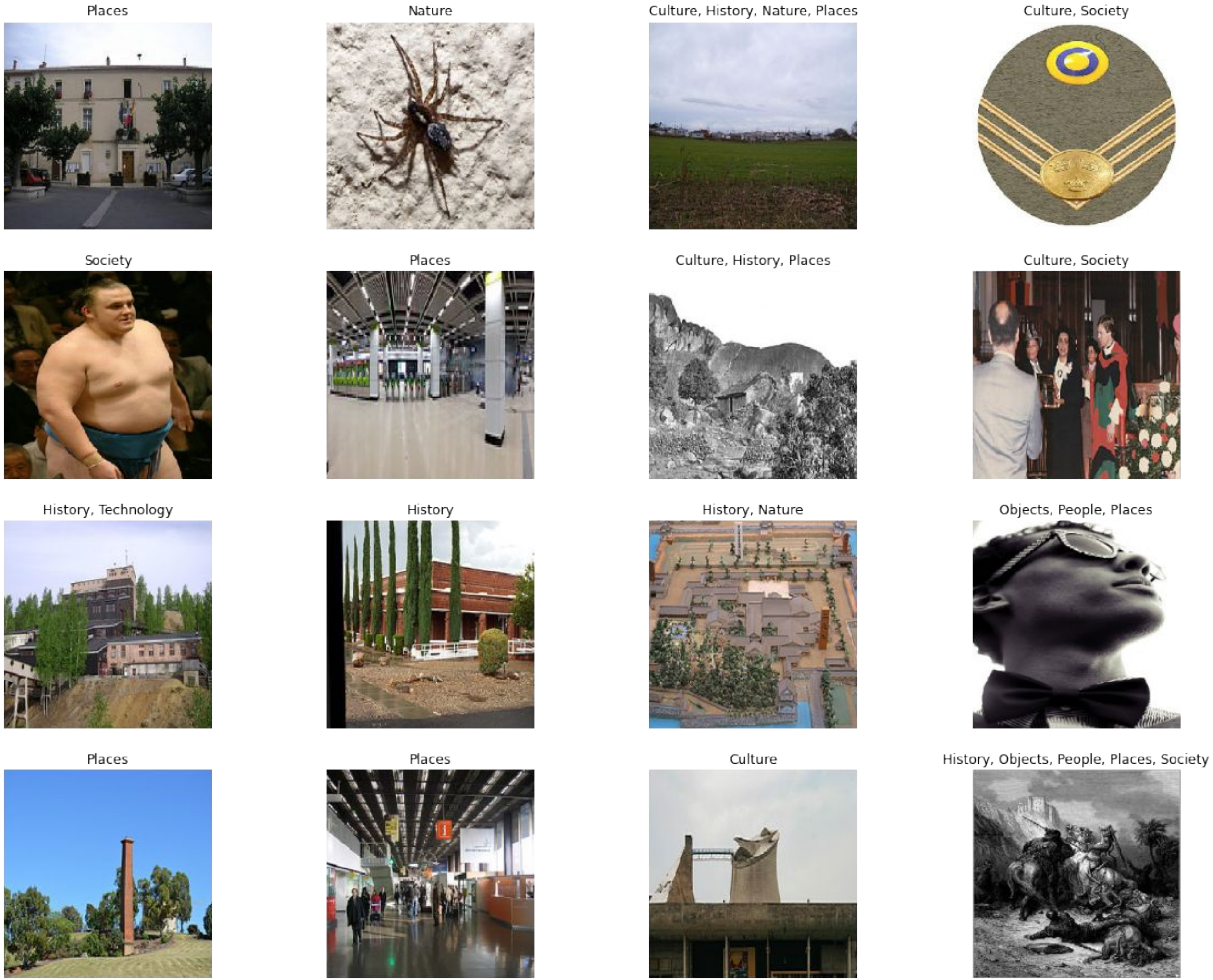
Images from WIT dataset

- Started with 11.5 unique images
- Restrictive filtering: 1.6M images used in training and testing



Matheus Bernat

Topic classification of Wikipedia images



Our work – data (labels)

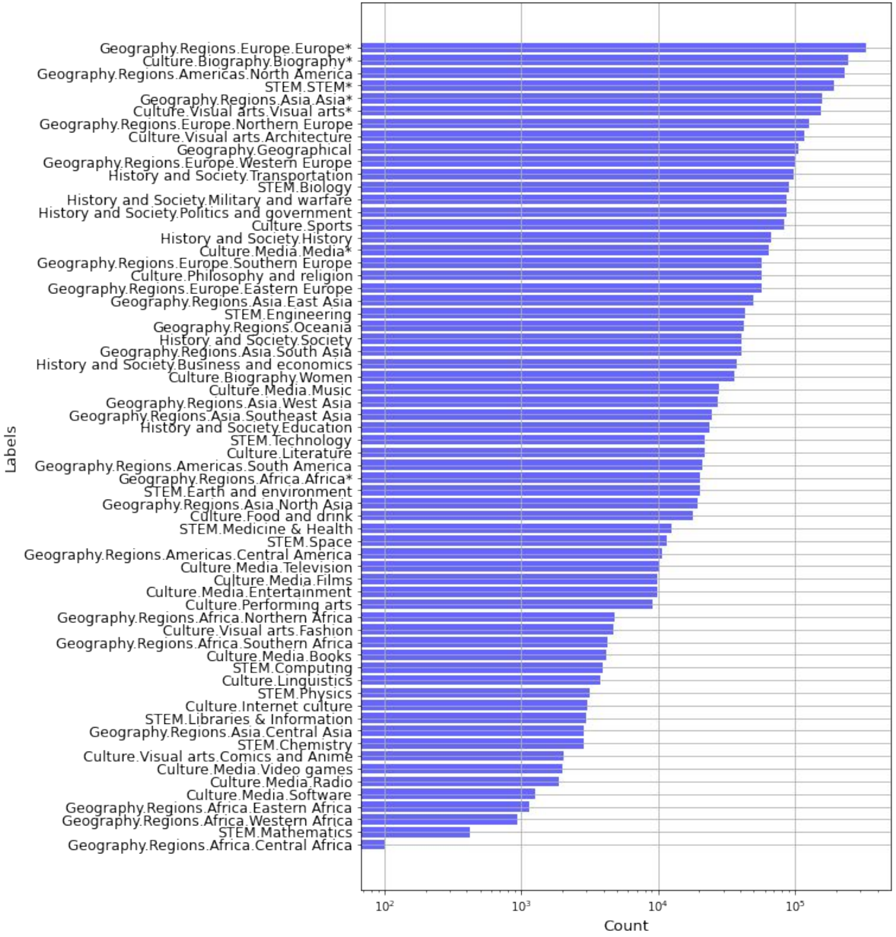
Image labels are either:

- the 64 ORES labels of all articles to which the images belong
- the 42 naive labels extracted by Salvi[7]

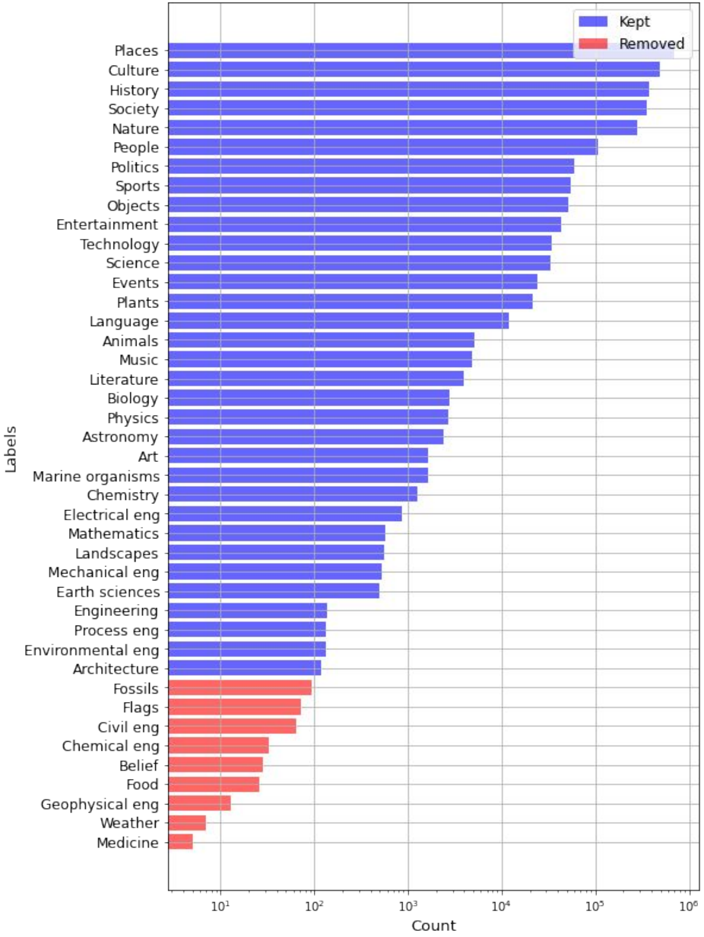
Heavily unbalanced label distribution!

EPFL

64 ORES labels



42 naive labels



Topic classification of Wikipedia images

Matheus Bernat

OUTLINE

EPFL

Experiment 1: ORES vs naive labels

- Same model (EfficientNetB0) and same images
- Difference: restricted data to have 10 of ORES, or naive labels

Table I
EVALUATION METRICS WHEN USING ORES LABELS.

	Precision	Recall	ROC AUC
Media	0.58	$\frac{429}{1360} = 0.32$	0.85
Music	0.64	$\frac{124}{614} = 0.20$	0.86
Sports	**0.87**	$\frac{700}{1790} = 0.39$	0.88
Visual arts	0.68	$\frac{1204}{3289} = 0.37$	0.84
Geographical	0.66	$\frac{509}{2267} = 0.23$	0.82
Military and warfare	0.64	$\frac{481}{1924} = 0.25$	0.81
Society	0.18	$\frac{7}{877} = 0.01$	0.66
Biology	0.80	$\frac{1138}{1939} = \mathbf{0.59}$	**0.93**
S.T.E.M.	0.81	$\frac{2126}{4203} = 0.51$	0.84
Space	0.85	$\frac{52}{254} = 0.21$	0.83
Micro average	0.74	$\boxed{0.37}$	0.87
Macro average	0.67	$\boxed{0.31}$	0.83

Table II
EVALUATION METRICS WHEN USING OUR CUSTOM LABELS.

	Precision	Recall	ROC AUC
Culture	0.64	$\frac{263}{9355} = 0.03$	0.62
Entertainment	0.21	$\frac{11}{795} = 0.01$	0.72
History	0.54	$\frac{511}{7216} = 0.07$	0.65
Nature	0.53	$\frac{1937}{5166} = 0.38$	0.77
Objects	0.16	$\frac{34}{937} = 0.04$	0.64
People	0.60	$\frac{35}{2042} = 0.02$	0.78
Places	**0.66**	$\frac{5558}{13288} = \mathbf{0.42}$	0.70
Politics	0.29	$\frac{158}{1074} = 0.15$	0.76
Society	0.52	$\frac{71}{6555} = 0.01$	0.65
Sports	0.45	$\frac{353}{1023} = 0.35$	**0.86**
Micro average	0.59	$\boxed{0.19}$	0.81
Macro average	0.46	$\boxed{0.15}$	0.72

Experiment 1: ORES vs naive labels

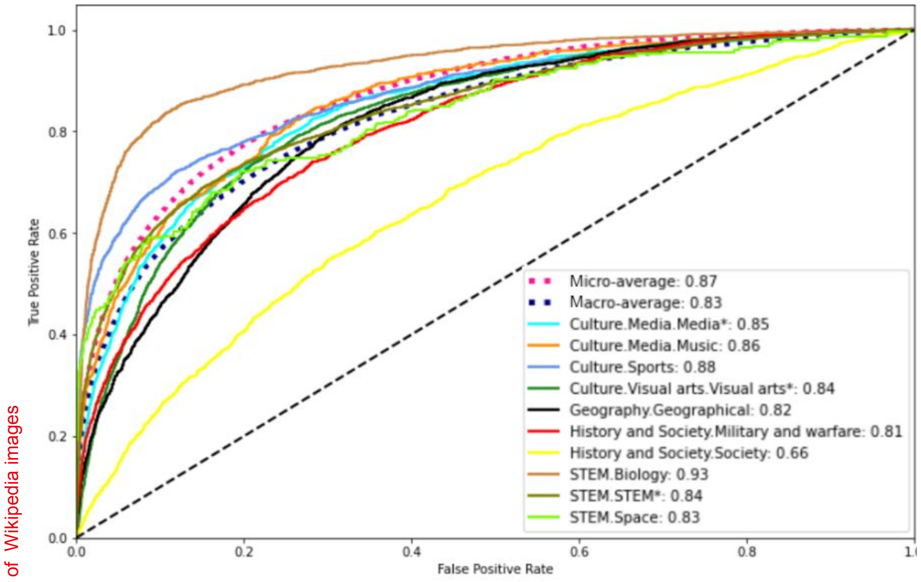


Figure 3. ROC curves for 10 **ORES labels**.

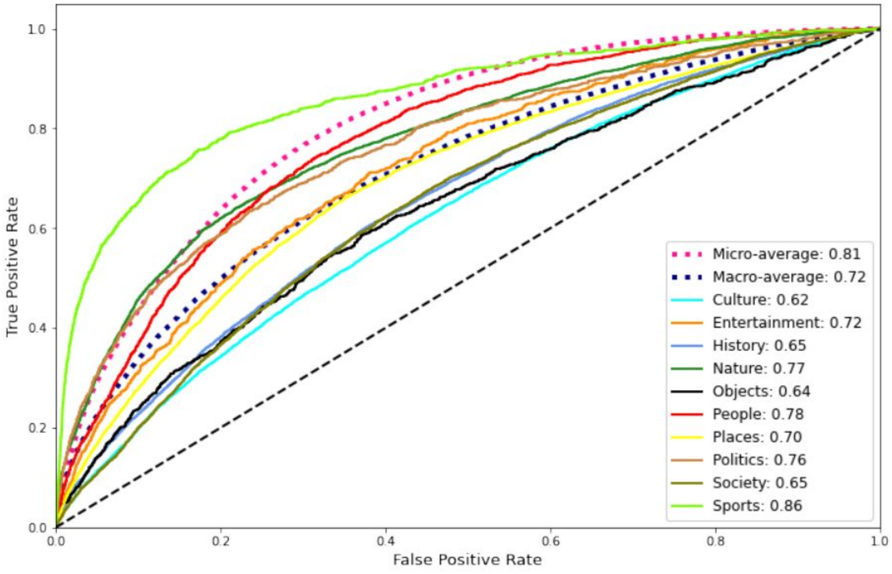


Figure 4. ROC curves for top 10 **custom labels**.

Experiment 2: EfficientNetB0 vs B2

- Same images and same 20 naive labels

Table III
EVALUATION METRICS FOR CUSTOM LABELS, 20 LABELS, EFFICIENTNETB0. 4.7M TOTAL PARAMETERS, 658K TRAINABLE PARAMETERS. MEAN NUMBER OF PREDICTED LABELS PER IMAGE: 0.11.

	Precision	Recall	ROC AUC
Animals	0.08	$\frac{42}{94} = 0.45$	**0.95**
Biology	0.03	$\frac{3}{49} = 0.06$	0.82
Culture	0.50	$\frac{1}{9355} = 0.00$	0.57
Entertainment	0.00	$\frac{0}{795} = 0.38$	0.70
Events	0.10	$\frac{5}{458} = 0.01$	0.61
History	1.00	$\frac{1}{7216} = 0.00$	0.53
Language	0.00	$\frac{0}{215} = 0.00$	0.73
Literature	0.00	$\frac{0}{81} = 0.00$	0.75
Music	0.07	$\frac{6}{85} = 0.07$	0.76
Nature	0.54	$\frac{105}{5166} = 0.02$	0.73
Objects	**1.00**	$\frac{1}{937} = 0.00$	0.59
People	0.44	$\frac{8}{2042} = 0.00$	0.76
Physics	0.00	$\frac{0}{35} = 0.00$	0.64
Places	0.71	$\frac{1134}{13288} = 0.09$	0.68
Plants	0.40	$\frac{177}{387} = \mathbf{0.46}$	**0.94**
Politics	0.37	$\frac{38}{1074} = 0.04$	0.73
Science	0.00	$\frac{0}{622} = 0.00$	0.60
Society	0.33	$\frac{1}{6555} = 0.00$	0.59
Sports	0.48	$\frac{131}{1023} = 0.13$	0.83
Technology	0.00	$\frac{1}{675} = 0.00$	0.56
Micro average	0.49	0.03	0.86
Macro average	0.30	0.04	0.70

Table IV
EVALUATION METRICS FOR CUSTOM LABELS, 20 LABELS, EFFICIENTNETB2. 8.5M TOTAL PARAMETERS, 723K TRAINABLE PARAMETERS. MEAN NUMBER OF PREDICTED LABELS PER IMAGE: 0.26.

	Precision	Recall	ROC AUC
Animals	0.09	$\frac{49}{94} = 0.52$	**0.96**
Biology	0.29	$\frac{2}{49} = 0.04$	0.80
Culture	0.00	$\frac{0}{9355} = 0.00$	0.55
Entertainment	0.00	$\frac{0}{795} = 0.38$	0.71
Events	0.06	$\frac{4}{458} = 0.01$	0.68
History	0.00	$\frac{0}{7216} = 0.00$	0.54
Language	0.00	$\frac{0}{215} = 0.00$	0.74
Literature	0.00	$\frac{0}{81} = 0.00$	0.80
Music	0.00	$\frac{0}{85} = 0.00$	0.79
Nature	0.49	$\frac{210}{5166} = 0.04$	0.74
Objects	0.00	$\frac{0}{937} = 0.00$	0.58
People	0.10	$\frac{4}{2042} = 0.00$	0.75
Physics	0.00	$\frac{0}{35} = 0.00$	0.63
Places	**0.67**	$\frac{3911}{13288} = 0.29$	0.68
Plants	0.40	$\frac{215}{387} = \mathbf{0.56}$	**0.95**
Politics	0.00	$\frac{0}{1074} = 0.00$	0.75
Science	0.00	$\frac{0}{622} = 0.00$	0.55
Society	0.00	$\frac{0}{6555} = 0.00$	0.60
Sports	0.53	$\frac{143}{1023} = 0.14$	0.85
Technology	0.13	$\frac{1}{675} = 0.00$	0.62
Micro average	0.59	0.19	0.86
Macro average	0.14	0.15	0.71

Experiment 2: EfficientNetB0 vs B2

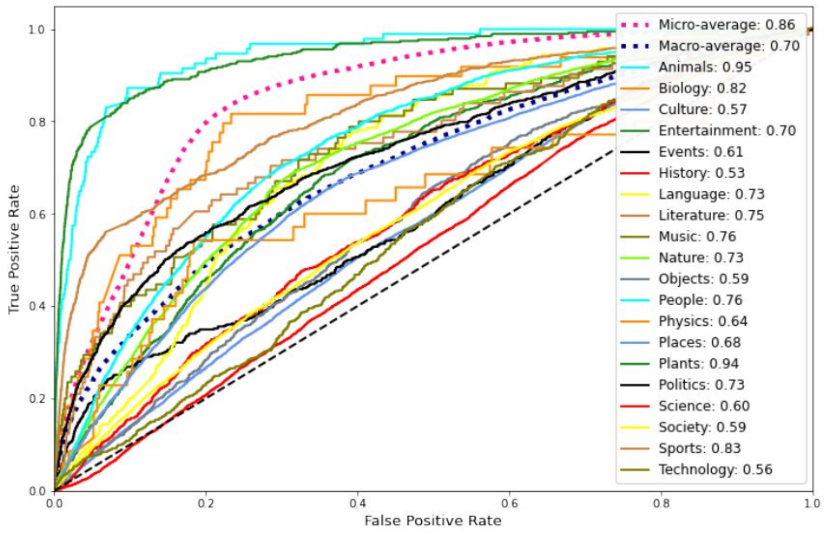


Figure 5. ROC curves for top 20 custom labels, **EfficentNetB0**-based model.

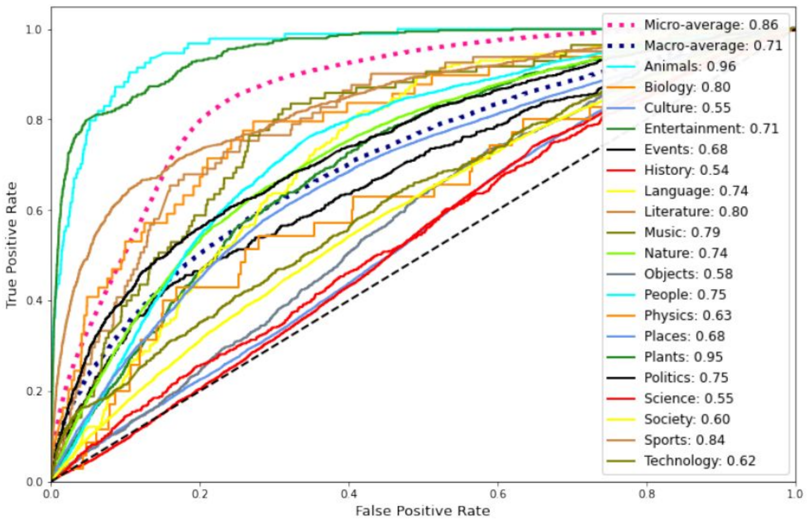


Figure 6. ROC curves for top 20 custom labels, **EfficentNetB2**-based model.

Main insights

1. The naive labels were inferior to the ORES labels according to our metrics.
2. The network with more parameters, EfficientNetB2, yielded higher prediction values (thus surpassing the 0.5 threshold more often) having greater recall, but does not outperform EfficientNetB0 significantly w.r.t ROC AUC.
3. The labels with better performance are those that are most present in the pre-training dataset (Plants and Animals)

References

[1] Miriam Redi. Discovering and Analyzing Wikipedia Images.
[2] K.D. Tran. Detecting Vandalism in Wikipedia in Multiple Languages. 2016
[3] M. Tan, Q. Le. EfficientNet: Rethinking Model Scaling for CNNs. 2020
[4] K. Srinivasan, K. Raman, J. Chen. WIT Dataset for Multimodal Multilingual ML. 2021
[5] A. Halfaker, R. S. Geiger. ORES: Lowering Barriers with Participatory ML in Wikipedia. 2019
[6] Miriam Redi. Prototypes of Image Classifiers Trained on Commons Categories. 2020

Thank you!