

```
In [1]: import matplotlib.pyplot as plt
import math

from IPython.display import IFrame
import ipyplot

In [2]: final_string = ""
for img in input("paste files ").split(" "):
    # if "pdf" in img:
    #     final_string += f'IFrame("{img}", width = "1152px", height = "580px")\n\n'
    # else:
    final_string += f'\n\n'
print(final_string)

paste files


In [37]: def grid(ims, lbls = None, img_width=500, show_url = False):
    if lbls is not None:
        ipyplot.plot_images(ims, labels = lbls, img_width=img_width, show_url=show_url)
    else:
        ipyplot.plot_images(ims, img_width=img_width, show_url=show_url)
```

Masters Thesis Colloquium

Proxy Attention : Approximating Attention in CNNs using Gradient Based Techniques

Subhaditya Mukherjee

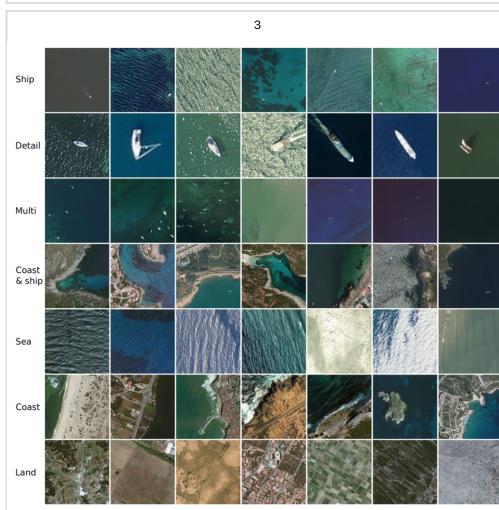
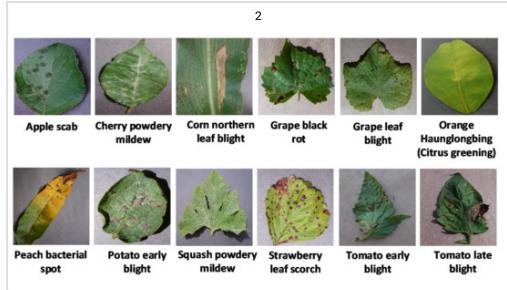
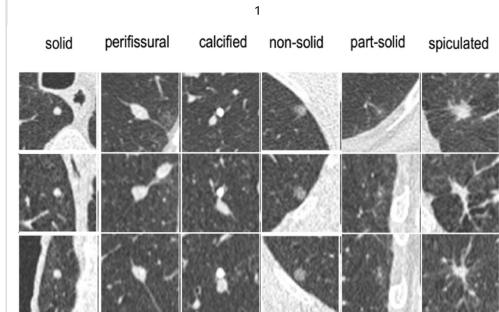
Supervisors: S.H. Mohades Kasaei and Matias Valdenegro

Image Classification

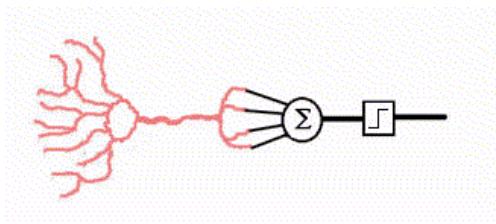
```
In [29]: ims = ["./images/cmuff.jpg", "./images/class2.png", "./images/class3.jpg", "./images/class4.png"]
```

```
In [30]: grid(ims)
```

```
show_html
```



How?



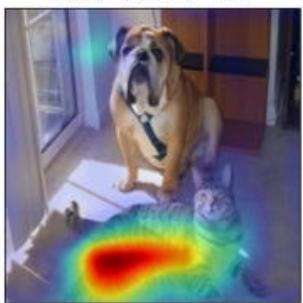
Quantifying Performance

Accuracy



Explainability

Grad-CAM for "Cat"

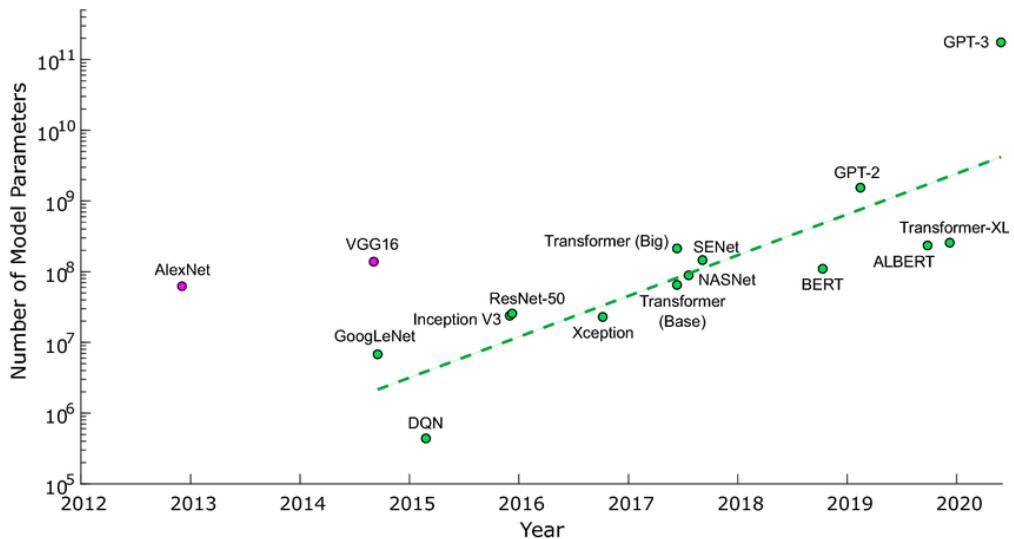


Grad-CAM for "Dog"

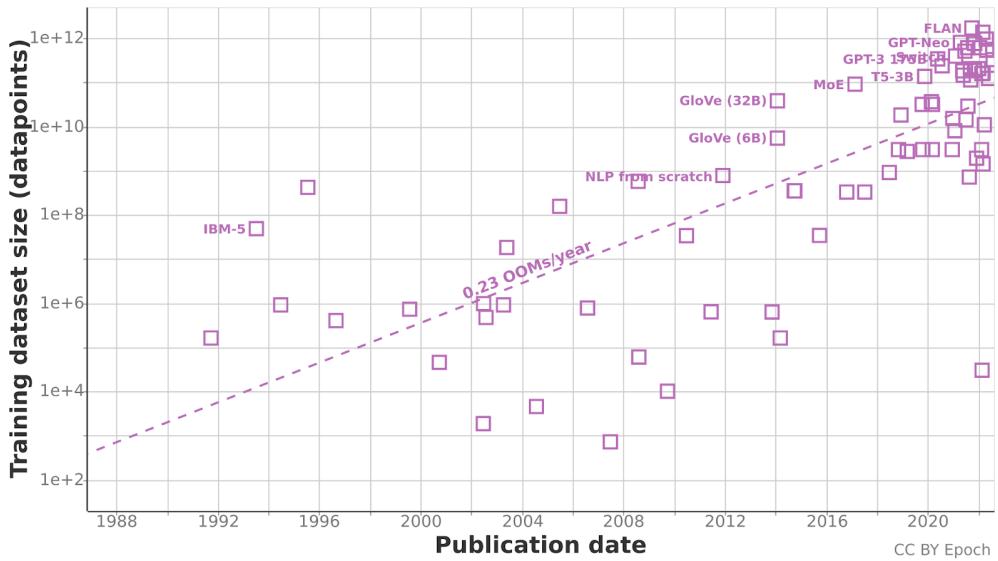


Challenges

Parameters



Dataset sizes



Consequences

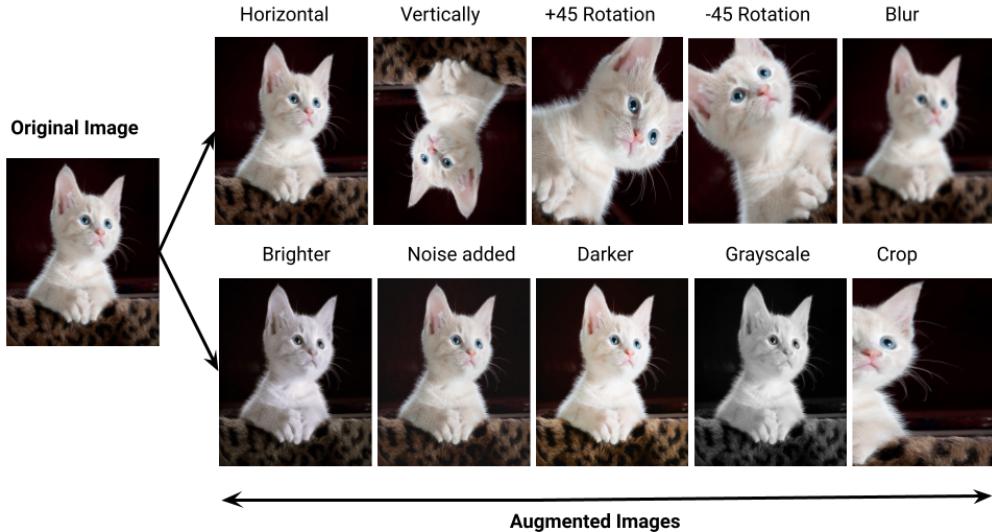
- More labelled data
- Vastly more energy consumption
- Funds

Objective

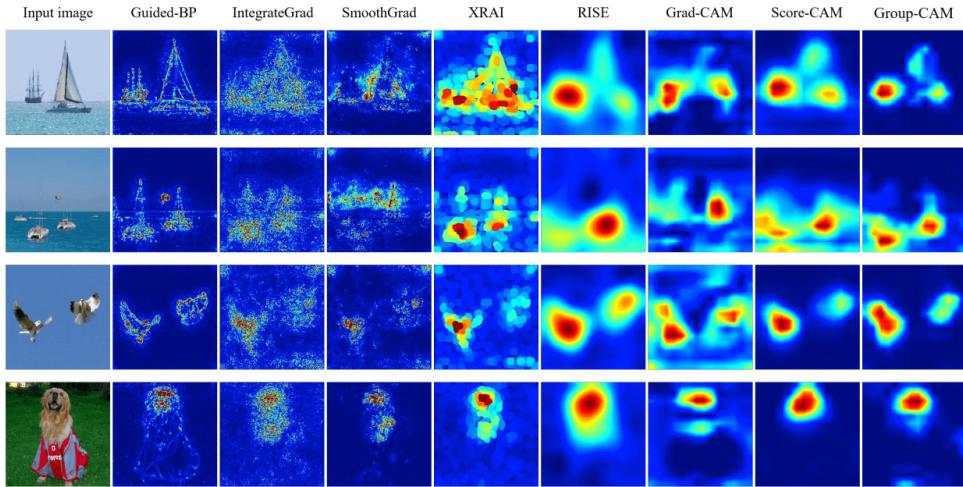
- Create a method to improve accuracy and explanations for image classification
- No extra labels, reduced compute time, no modification to the architecture

Previous Work

Augmentation



Gradient Based Explanations



Limitations of Previous Work

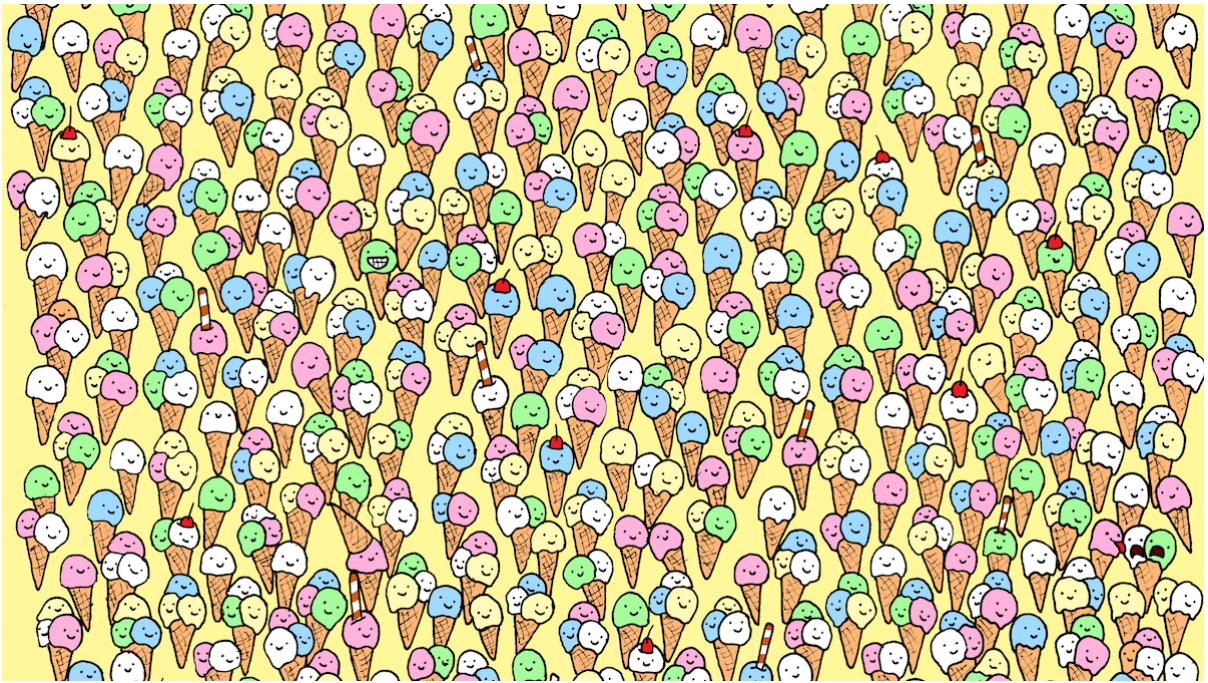
- Most of the algorithms are used as a final post-processing
- Contextual awareness
- Combining the fields of XAI and data augmentation to improve network performance is a rare practice.

Proxy Attention

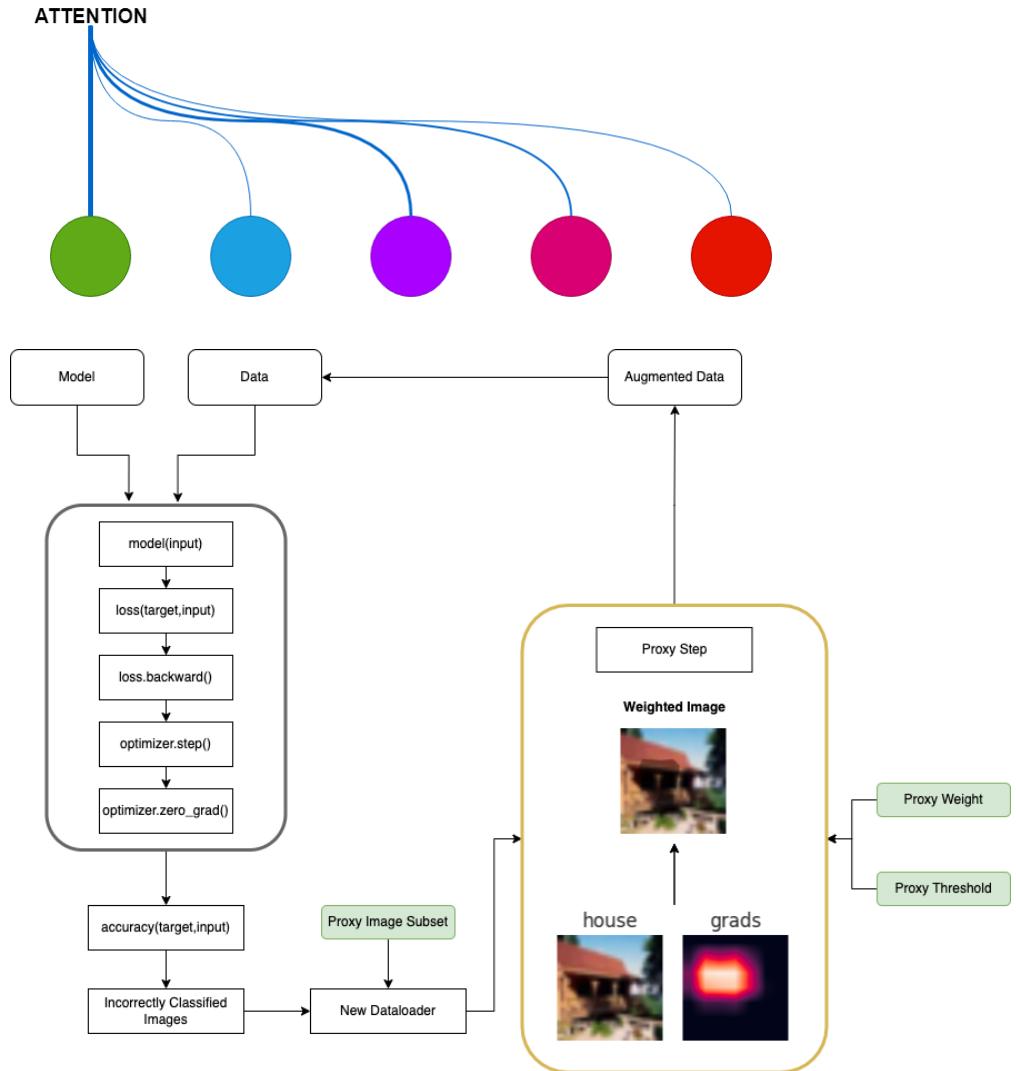
Research Questions

1. Is it possible to create an augmentation technique that uses Attention maps?
2. Is it possible to approximate the effects of Attention from ViTs in a CNN without changing the architecture?
3. Is it possible to make a network converge faster and consequently require fewer data using the outputs from XAI techniques?
4. Does using Proxy Attention impact the explainability positively?

Intuition



Backstage with the Vision Transformer

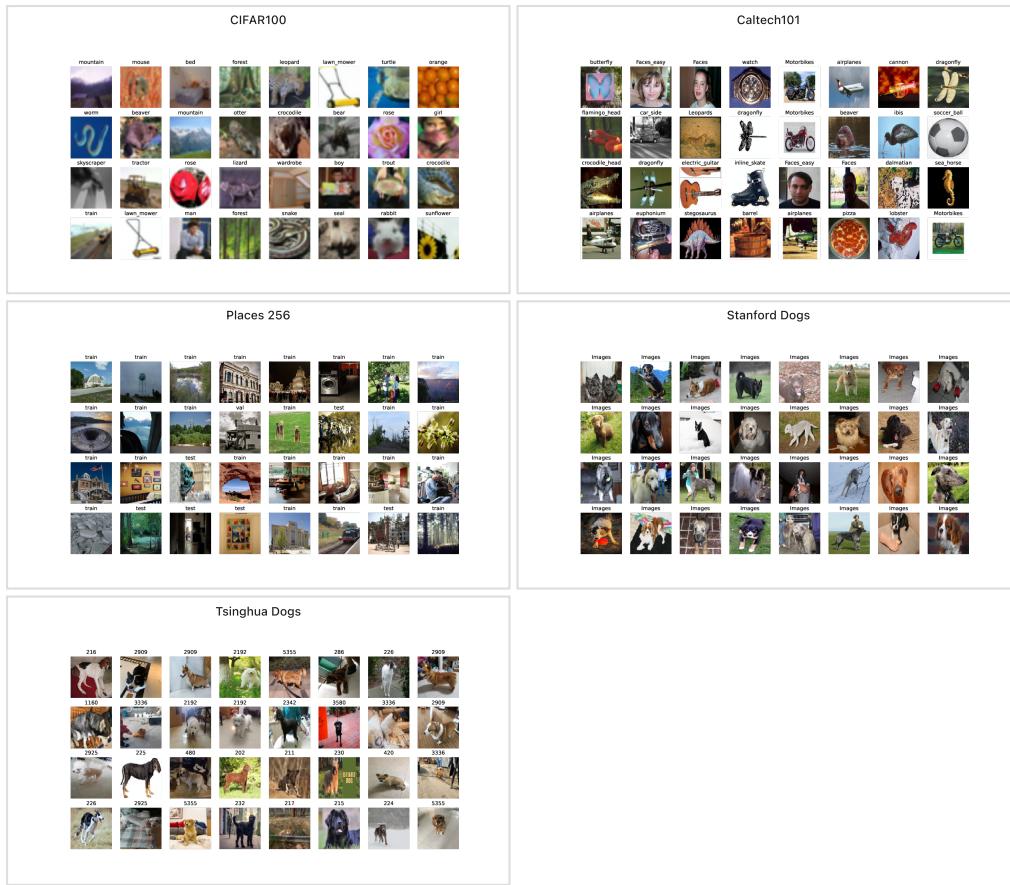


Testing

Datasets

```
In [41]: ims = ['./images/cifar100.pdf.png', './images/caltech101.pdf.png', './images/places256.pdf.png',
           './images/dogs.pdf.png', './images/tsing.png']
lbls = ['CIFAR100', 'Caltech101', 'Places 256', 'Stanford Dogs', 'Tsinghua Dogs']

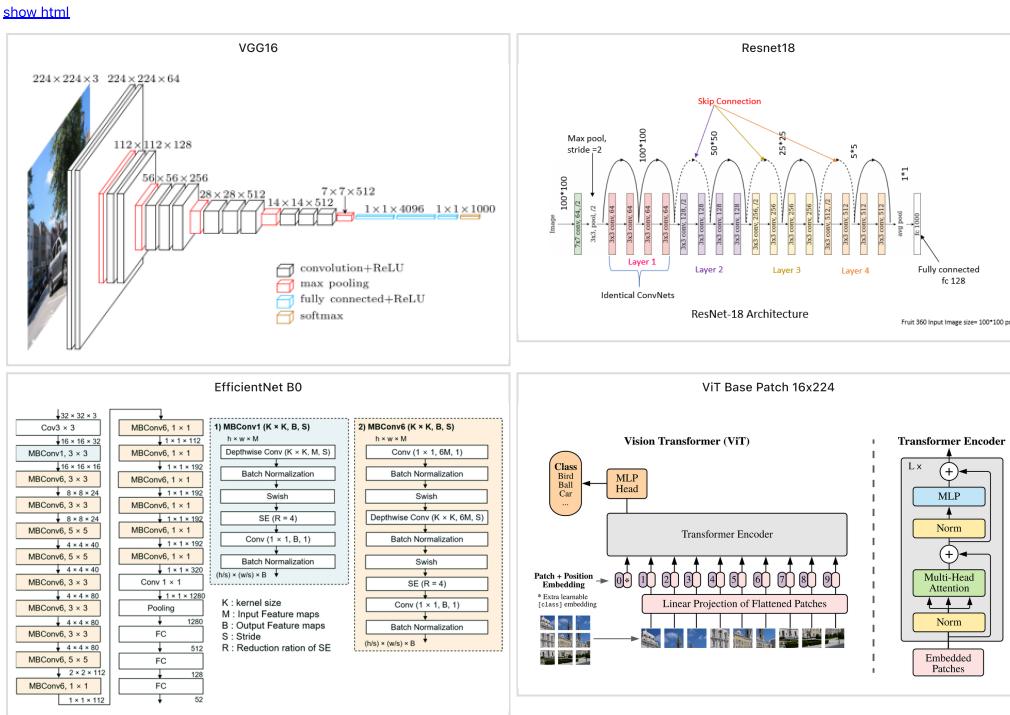
In [42]: grid(ims, lbls)
show.html
```



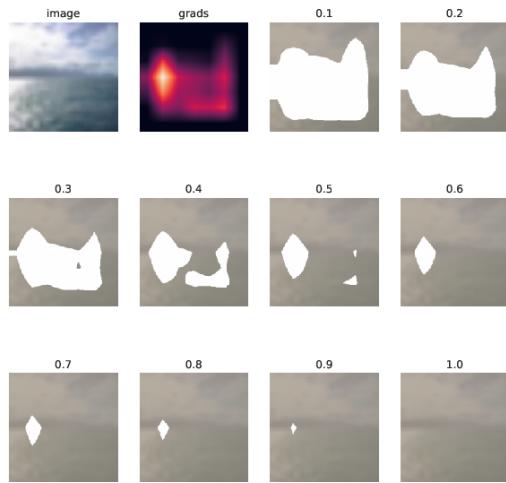
Architectures

```
In [43]: ims = ["./images/vggarch.png", "./images/resnetarch.png", "./images/effnetarch.png", "./images/vitarch.png"]
lbls = ["VGG16", "Resnet18", "EfficientNet B0", "ViT Base Patch 16x224"]
```

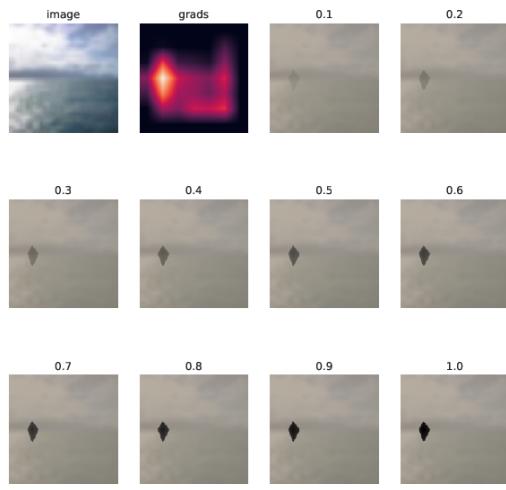
```
In [44]: grid(ims, lbls)
```



Proxy Image Threshold



Proxy Image Weight



Pixel Replacement Types



Proxy Step Schedule

- [20, p, 19]
- [5, p, 9, p, 9, p, 4]

Subset of Wrongly Classified Images

- 0.1
- 0.2
- 0.4
- 0.8
- 0.95

Training Assumptions

- Equal number of epochs
- Every other parameter fixed
- Equal number of data points for the DataLoaders

Results

By Dataset

```
In [46]: ims = ["./images/res1.png", "./images/res2.png", "./images/res3.png", "./images/res4.png"]
In [47]: grid(ims)
show.html
```

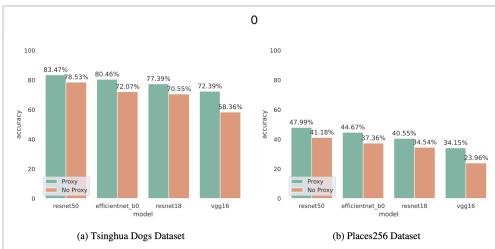


Figure 4.1: Comparing Accuracies of Models trained with and without Proxy Attention on the Tsinghua Dogs and Places256 datasets

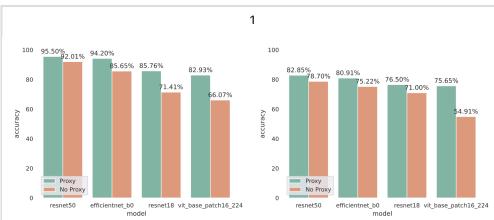


Figure 4.2: Comparing Accuracies of models trained with and without Proxy Attention on the Stanford Dogs and CIFAR100 datasets

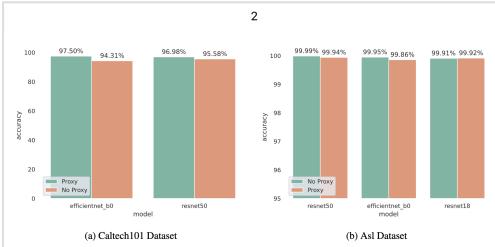


Figure 4.3: Comparing Accuracies of models trained with and without Proxy Attention on the Caltech101 and Asl datasets

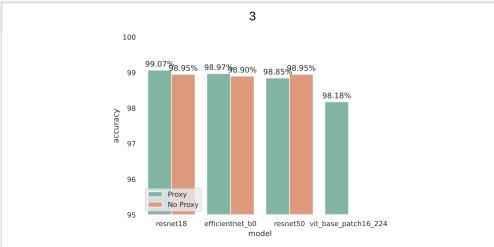


Figure 4.4: Comparing Accuracies of models trained with and without Proxy Attention on the Plant Disease dataset

By Hyper Parameters

```
In [48]: ims = ["./images/res5.png", "./images/res6.png", "./images/res7.png", "./images/res8.png"]
lbls = ["By Schedule", "By Proxy Threshold", "By Proxy Image Weight", "By Proxy Image Subset"]
```

```
In [49]: grid(ims, lbls)
```

[show html](#)

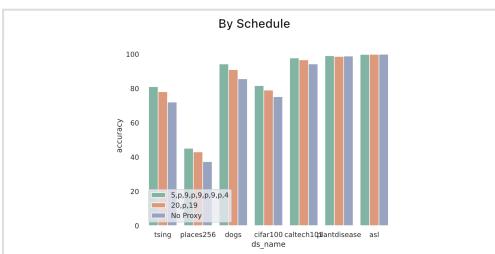


Figure 4.5: Comparing Accuracies of models trained with and without Proxy Attention on the ResNet50 [56] architecture for different step schedules

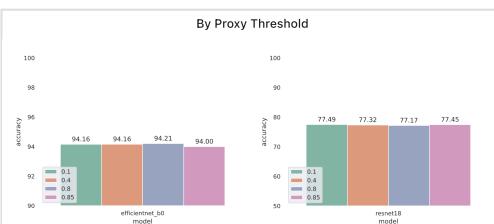


Figure 4.6: Comparing Accuracies of models trained with Proxy Attention for Different Proxy Thresholds

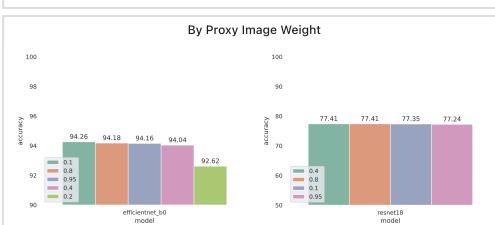


Figure 4.7: Comparing Accuracies of models trained with Proxy Attention for Different Proxy Image Weights

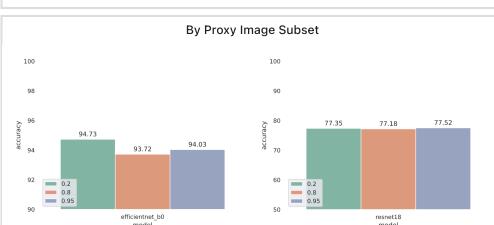


Figure 4.8: Comparing Accuracies of models trained with Proxy Attention for different Proxy Image Subsets

By Explainability

```
In [50]: ims = ["./images/expl2.png", "./images/expl3.png"]
```

```
In [52]: grid(ims, img_width=1000)
```

[show html](#)

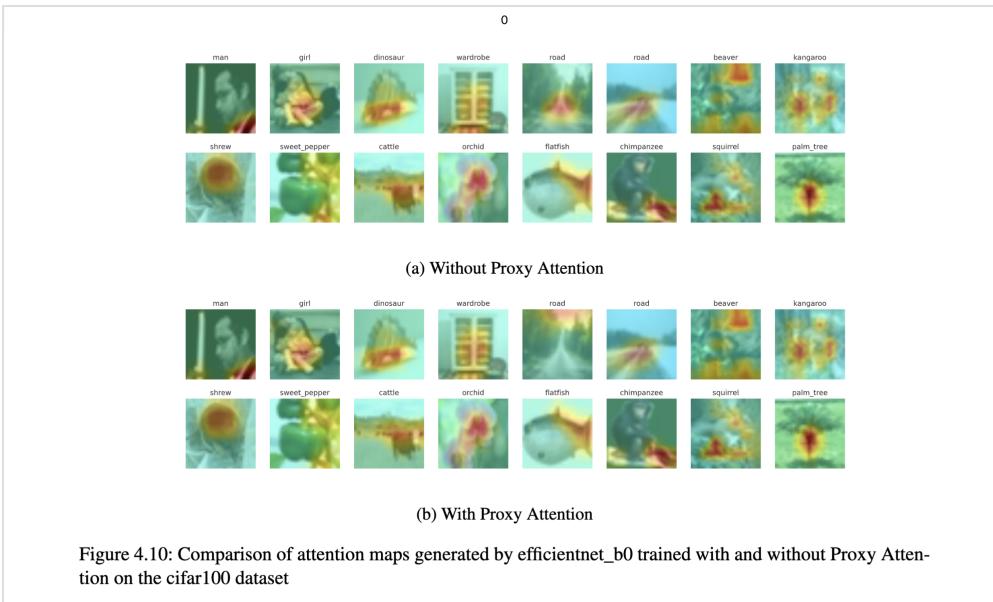


Figure 4.10: Comparison of attention maps generated by efficientnet_b0 trained with and without Proxy Attention on the cifar100 dataset

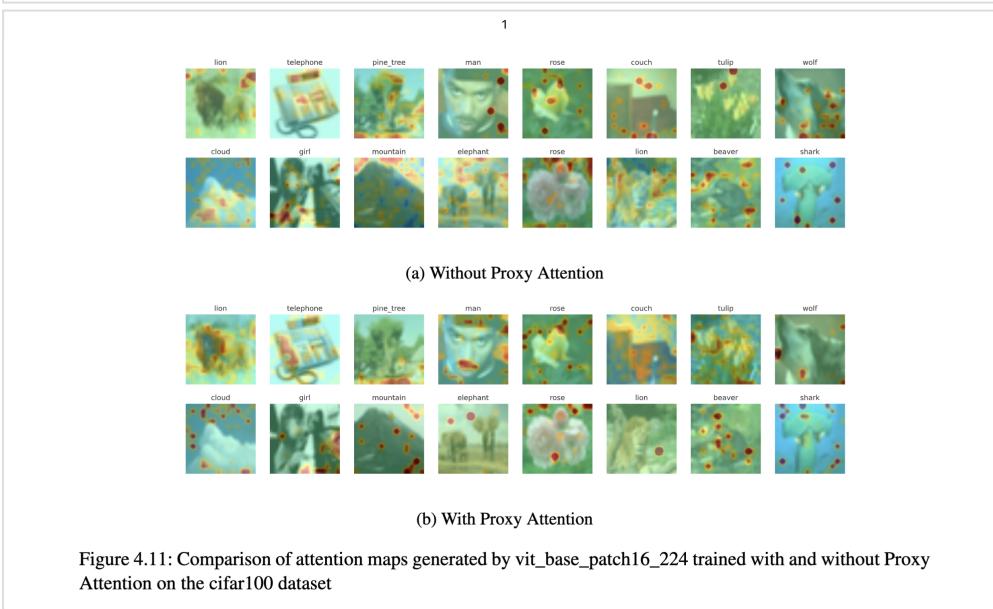


Figure 4.11: Comparison of attention maps generated by vit_base_patch16_224 trained with and without Proxy Attention on the cifar100 dataset

Discussion

- Improved performance
- Explainability
- When to apply Proxy Attention? : Easy vs Hard Datasets
- Optional hyperparameters : Proxy Weight and Proxy Threshold
- Scheduling the Proxy Step
- Performance across models

Limitations

- Hyperparameters
- Attention
- Better Scheduling

Future Work

- More Schedules
- More XAI methods
- Smoothing Attention Maps
- Better Attention Maps for ViT

Q&A



References

- <https://lih-verma.medium.com/query-key-and-value-in-attention-mechanism-3c3c6a2d4085>
- @dosovitskiyImageWorth16x162021
- <https://epochai.org/blog/trends-in-training-dataset-sizes>