Masking your problems away

Showing the effect of adding a masking layer on out of distribution accuracy

author: Quinten Nouwens supervisor: Wendelin Böhmer

1 The Problem

Motivation

Machine Learning solutions are deployed everywhere. For some of these applications correctness is important. These machine learning solutions are used in real life situations, which always change. Spurious correlations can make the machine learning solution perform poorly on data it has never seen before. It is therefore important that the effect of these correlations is minimized.

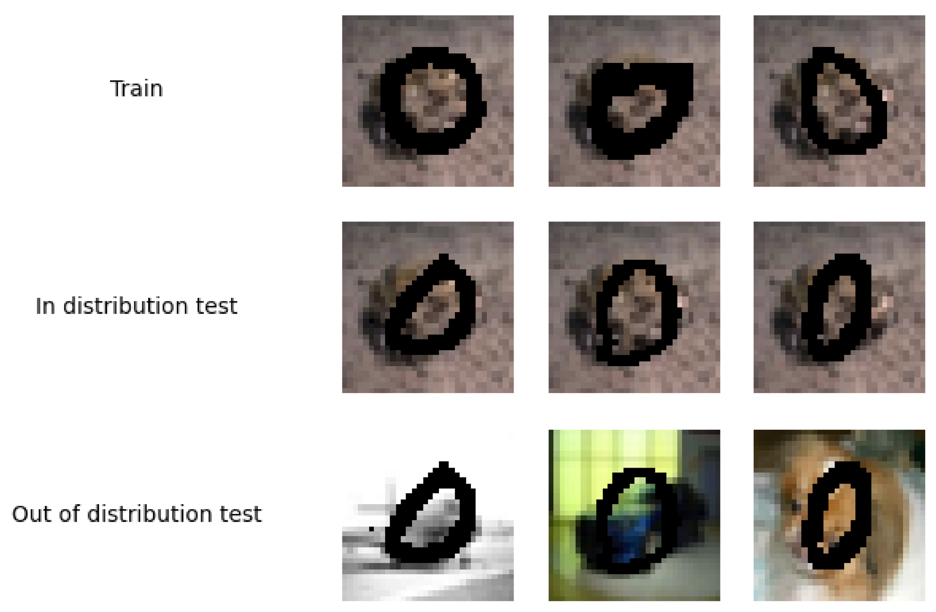

Figure: Example showing the difference between in distribution data and out of distribution data

Dataset

The dataset consists of MNIST digits overlaid on CIFAR-10 images, where during training only a certain number of backgrounds are used. These background are different for every digit. The amount of backgrounds per digit is called the class size.
The testing can be done both in distribution and out of distribution. In distribution testing is done on unseen MNIST samples overlaid on the same backgrounds as used during training. Out of distribution testing is done on unseen MNIST samples on top of unseen backgrounds. The accuracy is measured based on the ability to classify the digits.

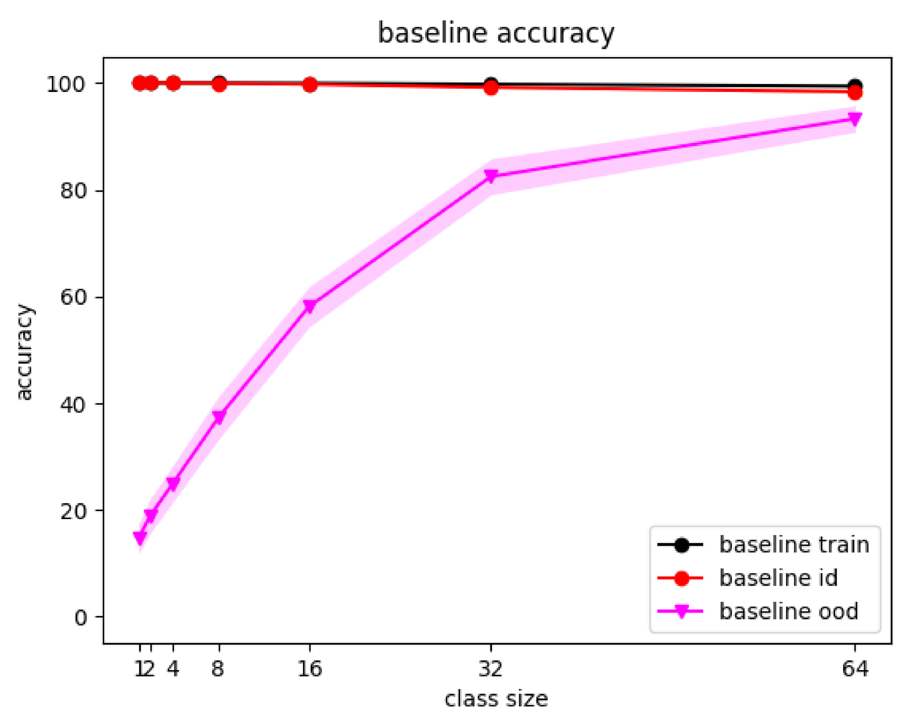

Figure: Difference between out of distribution accuracy and in distribution accuracy

Baseline

The figure above shows the baseline. This is the accuracy achieved by the LeNet-5 architecture. This architecture was chosen, because it has achieved a high accuracy on MNIST digits. It is also relatively small, which makes the effect clearer. From the baseline it can be seen that there is a gap between in distribution accuracy and out of distribution accuracy. Making this gap smaller is the aim of this research.

2 Noise

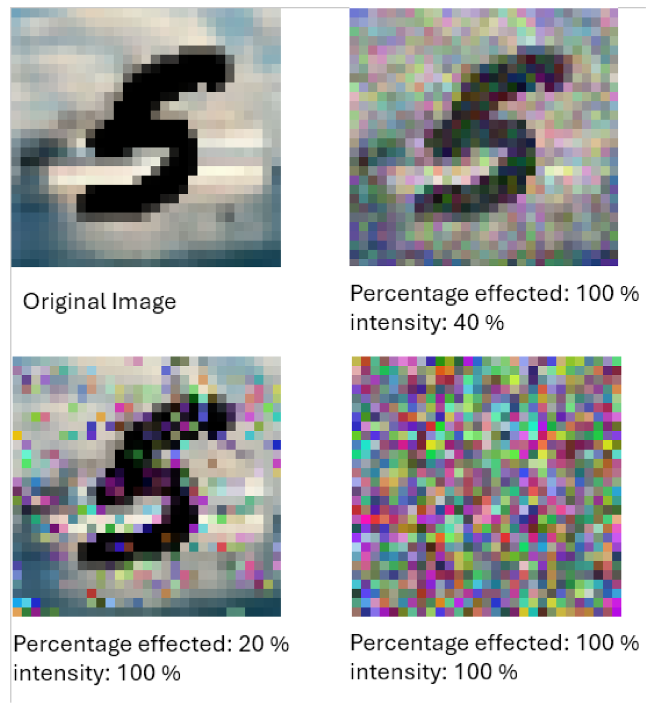

Figure: Noise application procedure

Noise application

For the noise results, noise is added to the full image during training. This is done using two parameters. Percentage affected and intensity of the noise. Percentage affected determines how many pixels of the image are affected by the noise. The intensity of the noise determines how strong the noise is.

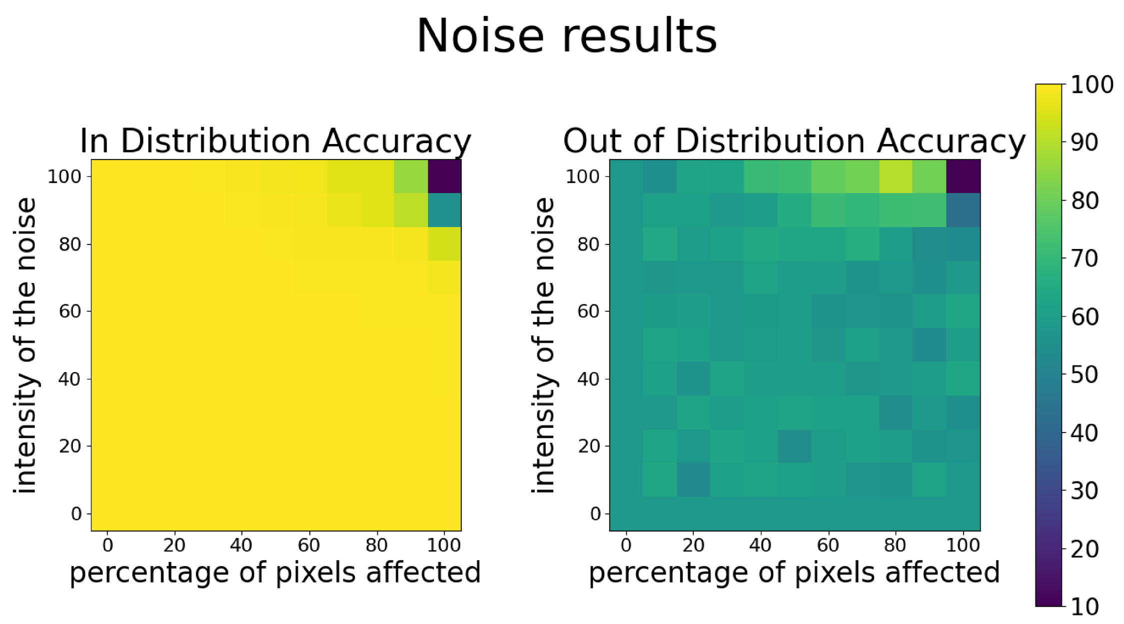

Figure: Mean accuracy achieved by models trained with noise for different percentage affected and different intensity of noise values. Obtained at a class size of 16

Noise results

The figure above shows that the in distribution accuracy is barely affected over the whole domain. Only when the image is only noise does the accuracy drop. The out of distribution accuracy is also fairly constant, except when the intensity of the noise reaches a 100%, at which point is is equivalent to dropout [2]. Due to this it was decided not to use noise for the masking solution.

3 Masker

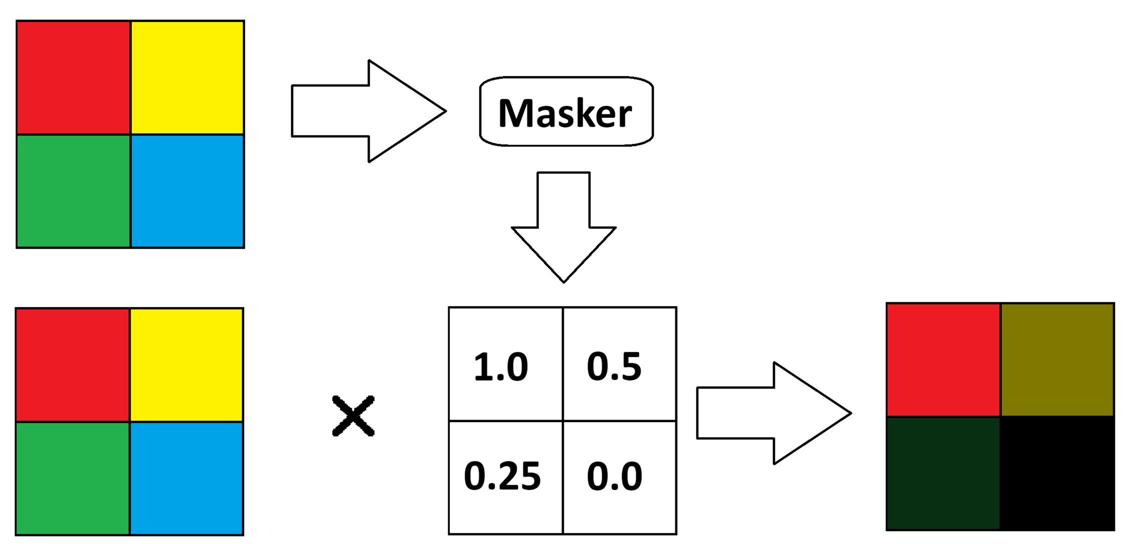

Figure: Example of mask application

Masking

Masking has been effective for complex visual tasks [1]. Masking is implemented by using a separate masking layer. This masking layer outputs a value between 0 and 1 for every pixel in the image. It then multiplies this mask with the original image.
For this project a constant was added before the image went through the masker. This was done because the masker without the constant can only make the image darker. This means that if the color of the MNIST sample is black, the masker can only make the background more similar to the foreground.

4 Results

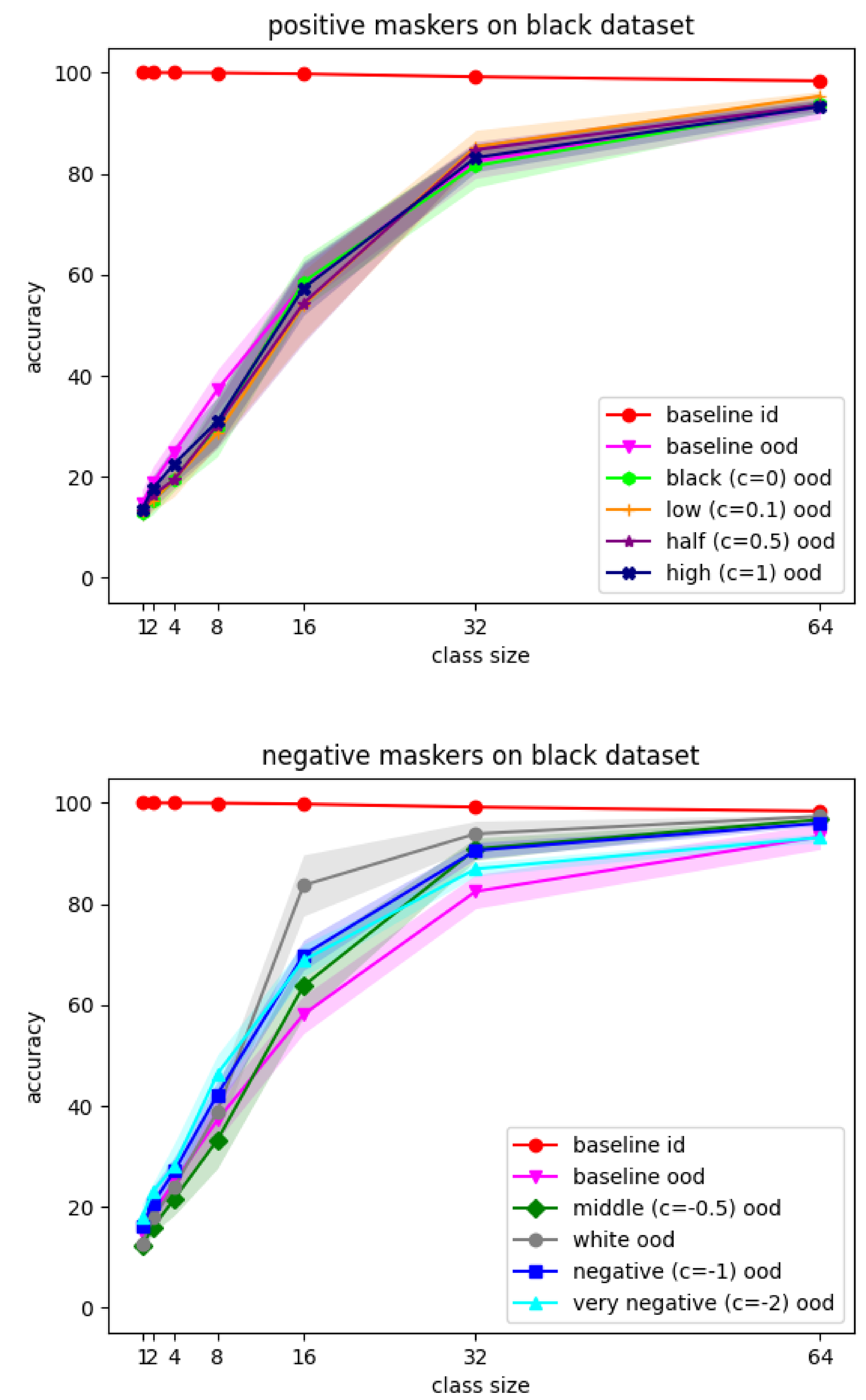

Figure: Results of maskers on out of distribution accuracy for the black dataset

Masker Results

The results above show that when a positive or no constant is added the accuracy does not increase with respect to the baseline accuracy. Maskers that have a negative constant do increase the out of distribution accuracy.

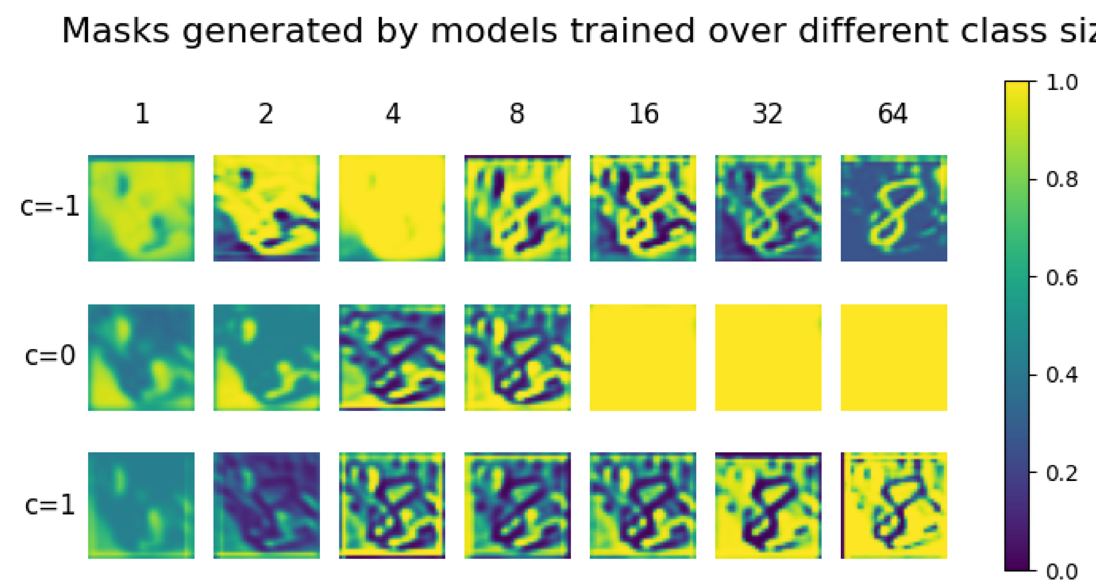

Figure: Masks generated by models trained over different class sizes and different added constants. The image used was out of distribution for all of them

Masks

The figure on the previous page shows masks generated by models, which were trained on a dataset with a class size that is shown above them and an added constant that is shown to the left of them. It shows a possible explanation for why the black masker (c=0) does not out perform the baseline. For class sizes higher than 8, the masks are all ones meaning just the original image is forwarded to the LeNet model.

Figure: Examples of different foregrounds on the same CIFAR sample and the same MNIST sample

Different Foregrounds

Because the results depended on which constant was used, it is interesting to see what happens when the foreground color is changed. As shown below, when a white foreground is used the results are flipped.

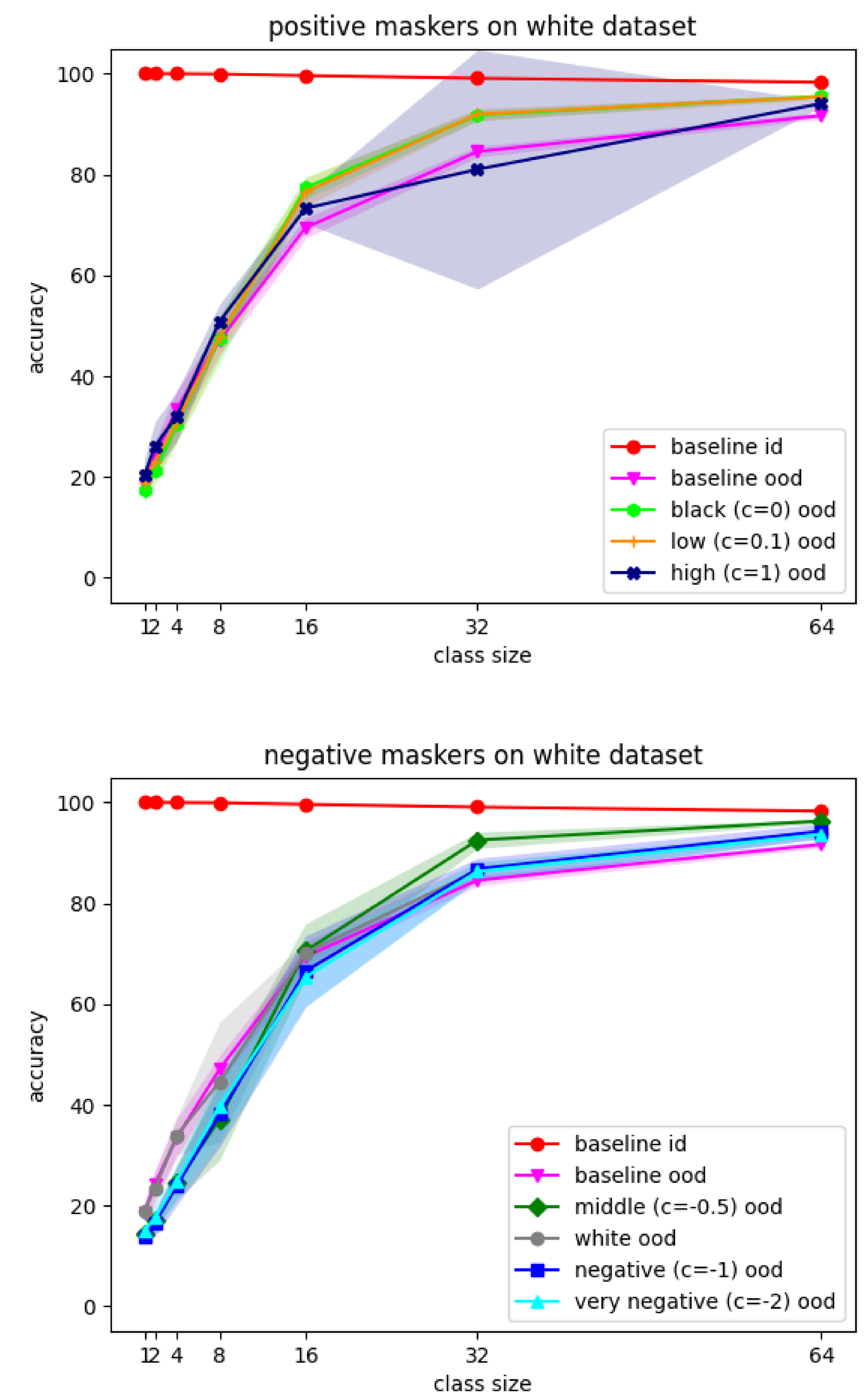

Figure: Results of maskers on out of distribution accuracy for the white dataset

5 conclusion

- Masks can be used to increase performance
- They are dataset dependent
- Noise does not increase generalization, except when it is equivalent to dropout.

References

[1] B. Grooten, T. Tomilin, G. Vasan, M. E. Taylor, A. R. Mahmood, M. Fang, M. Pechenizkiy, and D. C. Mocanu. *MaDi: Learning to Mask Distractions for Generalization in Visual Deep Reinforcement Learning*. 2023. DOI: 10.48550/arXiv.2312.15339.

[2] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958.

TU Delft
Delft University of Technology