Bag of Tricks for Training Brain-Like Deep Neural Networks

Alexander Riedel
Department of Industrial Engineering
Ernst-Abbe University for Applied Sciences
Jena, Germany
alexander.riedel@eah-jena.de

Abstract

The human-level performance of artificial neural networks (ANNs) in visual processing has made them a much-used research subject for understanding how the visual cortex really works. To assess how well various types of ANNs represent regions of the visual cortex, the Brain-Score platform provides several standardized benchmarks. These include the measure of explained variance in ventral stream regions V1, V2, V4, IT and the object recognition behavior in primates.
The aim of this work is to find a training procedure that maximizes an ANNs average score in the Brain-Score benchmark. The proposed pipeline combines a customized version of CutMix, heavy use of image augmentations, adversarial robust training, fixing the train-test resolution discrepancy, and weight averaging. Due to its widespread use, memory and computational efficiency, and object recognition performance, the EfficientNet-B1 architecture was used prototypically. The proposed training methods improve the public object recognition behavior metric score by 9% and the explained V1 variance by 62%, resulting in the best performing models in the Brain-Score competition 2022 . This is a strong indicator that finding the right training strategy might be crucial for developing brain-like ANNs.

1 Introduction

Deep artificial neural networks (ANN) have shown remarkable success in the field of object recognition and therefore became a popular subject of investigation in the neuroscience community. It was found that the latent features learned by ANNs after training on large image datasets resemble the neural responses in primates ventral visual cortices better than any other types of models [1, 2].

There is a large number of papers reporting how well ANNs account for various parts of the ventral stream such as V1, V2, V4, inferior temporal (IT) cortex or the object recognition behavior [1, 3, 4, 5, 6, 7]. However, many of these findings can hardly be compared across ANN models or different types of evaluation benchmarks. The Brain-Score platform provides a standardized framework for these kinds of benchmarks, including an API to (currently) perform 33 neuronal and behavioral brain-similarity tests on ANNs [8].

Many of the approaches in literature for finding brain-like ANNs focus on the search for beneficial ANN architectures rather than finding an optimal training and data augmentation strategy [6, 7, 9, 10]. This work is aiming towards finding a training procedure for established ANNs in order to maximize their average score in the Brain-Score benchmark. The focus of this work has been on training an ANN performing well on the V1 alignment benchmark by Freeman et al. [11] and the behavioral alignment benchmark by Rajalingham et al. [5]. Competitive results in the other evaluation categories V2, V4 and IT fell off with it.

1st Brain-Score Workshop (BSW 2022), Cascais, Portugal.

1.1 Related Work

The positive correlation between adversarial robustness and the ANNs capability to explain brain V1 variance has been demonstrated in literature [6, 12]. Dapello et. al. show that ANNs trained on adversarial perturbed images using a projected gradient descent (PGD) increase the ANNs V1 explained variance. Though the authors suggest a dedicated architectural backbone for increasing V1 explained variance and adversarial robustness of ANNs, the state-of-the-art method for training robust ANNs remains training on adversarial perturbed images [6, 13].

The behavior benchmark used in Brain-Score is established by Rajalingham et. al. and measures the ANNs similarity to primates in recognizing objects in different scenarios [5]. For determining the object recognition behavior, the test subjects are shown images of 3D object models in various poses on different backgrounds. Subsequently, the subjects are shown two possible objects (one previously shown and a distractor object) and asked to decide which one was shown in the previous composition. The benchmark then measures the decision-similarity between the ANN and the test subjects. To improve an ANNs behavior-similarity, a first intuitive approach would be to reduce the data distribution gap between an ANNs training data and the behavior test data.

The use of image augmentations in training ANNs, improves their generalization ability and thus prevent over-fitting the training data [14]. Augmentation techniques used in the proposed training pipeline include geometric transformations, color space augmentations, random erasing and adding noise.

For many ImageNet-pretrained vision ANNs, the current best training practice includes random cropping and resizing at training time, while simply extracting a centered square (center cropping) at test time [15]. This leads to different object sizes at train and test time, resulting in potentially wasted object recognition performance due to a distribution shift. The problem can be by increasing the resizing and cropping resolutions at test time.

Weight averaging (WA) is a well-known generalization technique in deep learning. In its simplest form, averaging the learned weights of two or more architectural identical ANNs increases the resulting ANN's performance. It also leads to wider optima than those found by stochastic gradient descent during training [16].

2 Methods and Results

The above-mentioned methods combine different concepts to improve an ANNs Brain-Score benchmarks. They represent a mix of regularization, generalization, data distribution shifts and adversarial robustness. Despite V1 alignment and adversarial robustness as well as behavioral alignment and classification accuracy generally don't correlate very strong, the robustness and accuracy gains from some techniques do correlate with the corresponding V1 and behavior metrics. Fixing the train-test resolution discrepancy (FixRes) of EfficientNet-B1 for example changes an ANNs Brain-Score behavior metric in the same cyclic pattern as observed for the object classification accuracy (see Figure 1 and [17]). Also, the behavior score gains from weight averaging are comparable to the image classification accuracy gains from this method. To evaluate the benefit of the proposed training and evaluation pipeline, an ablation study was conducted. The study is split into behavior metric ablation and V1 metric ablation, both using different techniques.

An intuitive approach to train ANNs to be not distracted by the background image presented in the behavioral benchmark [5] would be a adjusted form of the CutMix augmentation [18]. In contrast to the original CutMix, the proposed method only accounts for the pasted image to be the correct class (Table 3, "Foreground-CutMix"). This technique will shift the training data towards the data shown in the behavioral benchmark.

The augmentations used in the ablation study (referred to as "heavy augmentations") include image shifting, rotation, elastic deformation, brightness and contrast adjustments, motion blurring, cutout, adding Gaussian and ISO noise [19].

The adversarial robust training used for improving the V1 metric adversely effects the ANNs object classification performance, and thus the behavior score. To overcome this, the final ANN was first trained adversarial robust under the optimal parameters (see Table 2). Subsequently the first half of the ANNs layers were frozen (including Batch Norm statistics), and it was fine-tuned for behavior response using the proposed techniques in 1. The layers that explain most of V1 variance are likely to be found among the first half of all layers. The behavioral response however is read out in the penultimate layer. The presented ablation studies were conducted without any layer freezing.

Table 1: EfficientNet-B1 fine-tuned for 5 epochs using the ADAM optimizer and an initial learning rate of 1.5e-4.

Public Behavior (Rajalingham2018-i2n)	Heavy Aug	CutMix	F-CutMix	FixRes	WA
0.5133					
0.5281	✓				
0.5129		✓			
0.5200			✓		
0.5365	✓		✓		
0.5429	✓		✓	✓	
0.5563	✓		✓	✓	✓

The frozen layer fine-tuning for behavioral response leads to a lower behavior score compared to all-layer training due to the limited capacity. However, the proposed training techniques boost the public behavior score from 0.4527 ($L_\infty, \epsilon = 8/255$ adversarially trained) to 0.5271 ($L_\infty, \epsilon = 8/255$ adversarially trained and behavior fine-tuned) while keeping the optimized V1 explained variance.

The proposed methods were validated on a ResNet50 [20] to demonstrate their generalizability. Similar improvements in explained V1 variance and behavioral response can be achieved compared to the studied EfficientNet-B1 architecture.

Table 2: EfficientNet-B1 trained adversarially robust [21] under different constraints and epsilon values were evaluated on their V1 explained variance using the Freeman-Ziemba 2013.V1-pls score. The ANN validated without robust training is the best performing behavior ANN (see Table 1).

.

Constraint	Public V1 (Freeman-Ziemba 2013.V1-pls)
w/o robust training	0.2257
$L_2, \epsilon = 3.0$	0.3302
$L_\infty, \epsilon = 8/255$	**0.3661**
$L_\infty, \epsilon = 4/255$	0.3401

3 Discussion

The combination of adversarial robust training, layer freezing and behavioral metric aware data distribution shifting leads to an improved explained V1 variance and a competitive behavioral metric response. Despite the proposed methods lead to a winning model in the Brain-Score competition 2022 and present a new state-of-the-art performance in the Brain-Score benchmark overall, it still remains unclear whether they lead towards biologically more plausible ANNs or if they are just ways to exploit the Brain-Score benchmark. However, the further development of brain-like ANNs should address the influence of an optimized training pipeline as it might be a way to push a biologically plausible ANN towards its frontiers of explaining primates visual cortex responses.

References

[1] D. L. K. Yamins et al. "Performance-optimized hierarchical models predict neural responses in higher visual cortex". In: *Proceedings of the National Academy of Sciences* 111.23 (June 2014), pp. 8619–8624. ISSN: 0027-8424. DOI: 10.1073/pnas.1403112111. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1403112111.

[2] Nikolaus Kriegeskorte. “Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing”. In: *Annual Review of Vision Science* 1.1 (Nov. 2015), pp. 417–446. ISSN: 2374-4642. DOI: 10.1146/annurev-vision-082114-035447. URL: https://www.annualreviews.org/doi/10.1146/annurev-vision-082114-035447.

[3] U. Guclu and M. A. J. van Gerven. “Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream”. In: *Journal of Neuroscience* 35.27 (July 2015), pp. 10005–10014. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.5023-14.2015. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.5023-14.2015.

[4] Chengxu Zhuang et al. “Unsupervised neural network models of the ventral visual stream”. In: *Proceedings of the National Academy of Sciences* 118.3 (Jan. 2021), e2014196118. ISSN: 0027-8424. DOI: 10.1073/pnas.2014196118. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.2014196118.

[5] Rishi Rajalingham et al. “Large-Scale, High-Resolution Comparison of the Core Visual Object Recognition Behavior of Humans, Monkeys, and State-of-the-Art Deep Artificial Neural Networks”. In: *The Journal of Neuroscience* 38.33 (Aug. 2018), pp. 7255–7269. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.0388-18.2018. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0388-18.2018.

[6] Joel Dapello et al. “Simulating a Primary Visual Cortex at the Front of CNNs Improves Robustness to Image Perturbations”. In: *bioRxiv* (Jan. 2020), p. 2020.06.16.154542. DOI: 10.1101/2020.06.16.154542. URL: http://biorxiv.org/content/early/2020/10/22/2020.06.16.154542.abstract.

[7] Jonas Kubilius et al. “Brain-like Object Recognition with High-Performing Shallow Recurrent ANNs”. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2019.

[8] Martin Schrimpf et al. “Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?” In: *bioRxiv* (Jan. 2018), p. 407007. DOI: 10.1101/407007. URL: http://biorxiv.org/content/early/2018/09/05/407007.abstract.

[9] Anne-Ruth José Meijer and Arnoud Visser. “A shallow residual neural network to predict the visual cortex response”. In: (June 2019). arXiv: 1906.11578. URL: http://arxiv.org/abs/1906.11578.

[10] Roman Pogodin et al. “Towards Biologically Plausible Convolutional Networks”. In: (June 2021). arXiv: 2106.13031. URL: http://arxiv.org/abs/2106.13031.

[11] Jeremy Freeman et al. “A functional and perceptual signature of the second visual area in primates”. In: *Nature neuroscience* 16 (2013), pp. 974–981.

[12] Zhe Li et al. “Learning From Brains How to Regularize Machines”. In: (Nov. 2019). arXiv: 1911.05072. URL: http://arxiv.org/abs/1911.05072.

[13] Aleksander Madry et al. “Towards Deep Learning Models Resistant to Adversarial Attacks”. In: (June 2017). arXiv: 1706.06083. URL: http://arxiv.org/abs/1706.06083.

[14] Connor Shorten and Taghi M. Khoshgoftaar. “A survey on Image Data Augmentation for Deep Learning”. In: *J. Big Data* 6 (2019), p. 60. DOI: 10.1186/s40537-019-0197-0. URL: https://doi.org/10.1186/s40537-019-0197-0.

[15] Hugo Touvron et al. *Fixing the train-test resolution discrepancy*. 2022. arXiv: 1906.06423 [cs.CV].

[16] Pavel Izmailov et al. “Averaging weights leads to wider optima and better generalization”. English (US). In: *34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018*. Ed. by Ricardo Silva, Amir Globerson, and Amir Globerson. 34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018. Association For Uncertainty in Artificial Intelligence (AUAI), 2018, pp. 876–885.

[17] Hugo Touvron et al. *Fixing the train-test resolution discrepancy: FixEfficientNet*. 2020. arXiv: 2003.08237 [cs.CV].

[18] Sangdoo Yun et al. *CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features*. 2019. arXiv: 1905.04899 [cs.CV].

[19] Alexander Buslaev et al. “Albumentations: Fast and Flexible Image Augmentations”. In: *Information* 11.2 (2020). ISSN: 2078-2489. DOI: 10.3390/info11020125. URL: https://www.mdpi.com/2078-2489/11/2/125.

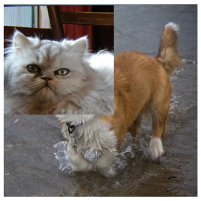
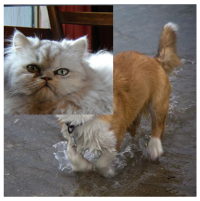
[20] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[21] Logan Engstrom et al. *Robustness (Python Library)*. 2019. URL: https://github.com/MadryLab/robustness.

A Appendix

As the images presented in the Rajalingham behavior benchmark are composed of an 3D objects randomly pasted on different backgrounds, it seemed advantageous to simulate this type of input. An intuitive approach is simply pasting parts of ImageNet images on top of a background image and assign the pasted image label to the whole image. This method is called Foreground-CutMix in the present work.

Table 3: Foreground-CutMix assigns the label of a randomly pasted image part to the whole image. This induces the concept that an object can be anywhere in the image and shifts away from photographers biases.

	Cutmix	Foreground-CutMix
Image		
Label	Dog 0.6 Cat 0.4	Cat 1.0

The train-test resolution discrepancy of ANNs that were trained using random resizing and cropping and tested using center cropping leads to a decrease in classification accuracy. This can be overcome by manipulating the final average pooling layer or by simply finding the best performing test resolution. For an EfficientNet-B1 trained on images that are resized to 272x272 pixels and cropped to 240x240 pixels, the behavior-benchmark-optimal resolution at test time was empirically determined to be 324x324 pixels. A similar effect is observed for the ImageNet object classification accuracy in [15], though the optimal EfficientNet-B1 resolution is found to be 384x384 pixels for object classification.

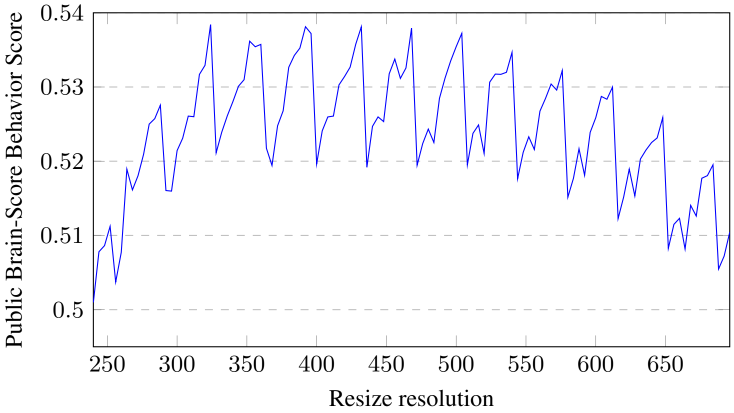


Figure 1: The Brain-Score behavior score is dependent on the input image size for ANNs that are trained using random crop resizing. A similar pattern is observed for the object classification performance [15].