Equiangular Basis Vectors (Supplementary Materials)

Yang Shen Xuhao Sun Xiu-Shen Wei*
School of Computer Science and Engineering, Nanjing University of Science and Technology, China
{shenyang_98,sunxh,weixs}@njust.edu.cn

In the supplementary materials, we provide more experiments for the proposed Equiangular Basis Vectors (EBVs).

1. Relations between α, d and N

In Section 3.1 of the paper, we make the definition of the proposed EBVs, where $\alpha \in [0, 1)$ represents the maximum value of the absolute value of the cosine of the angle between any two vectors, d denotes the dimension of each coordinate vector while N denotes the number of the basis vectors. Specifically, for the EBVs set $\mathcal{W}$, each $\boldsymbol{w} \in \mathbb{R}^d$ in $\mathcal{W}$ should satisfies:

$$\forall \boldsymbol{w}_i, \boldsymbol{w}_j \in \mathcal{W}, i \neq j, \quad -\alpha \leq \frac{\boldsymbol{w}_i \cdot \boldsymbol{w}_j}{\|\boldsymbol{w}_i\| \, \|\boldsymbol{w}_j\|} \leq \alpha \,, \quad (1)$$

where $\|\cdot\|$ denotes the Euclidean norm and $\mathrm{card}(\mathcal{W}) = N$.

According to Elad et al. [3], we have known that we can construct a Grassmannian matrix if N satisfies:

$$N < \min(d(d+1)/2, (N-d)(N-d+1)/2) \,, \quad (2)$$

while the lower bound for α equals $\sqrt{\frac{N-d}{d(N-1)}}$. Therefore, we could get a set $\mathcal{W}'$ ($\mathrm{card}(\mathcal{W}') = N$) which satisfies:

$$\forall \boldsymbol{w}_i, \boldsymbol{w}_j \in \mathcal{W}', i \neq j, \quad 0 \leq \frac{\boldsymbol{w}_i \cdot \boldsymbol{w}_j}{\|\boldsymbol{w}_i\| \, \|\boldsymbol{w}_j\|} \leq \alpha \,. \quad (3)$$

However, if N does not satisfy Eq. (2) or the fixed α is larger than the lower bound, we can not construct such a Grassmannian matrix. Furthermore, we would like to explore the relations between α, d and N. Thus, we use the bisection method to search for the maximum value of N when given fixed α and d which satisfies Eq. (1) according to Algorithm 1 in the paper. In Figure 1 in the supplementary materials, we draw the relationship curve between α, d and N. Specifically, when fixed α and d, we calculate a progressive upper bound for N. Additionally, it can be easily proved that we can find n ($2 \leq n \leq N$) vectors which satisfy Eq. (1) when given the same α and d.

2. Empirical evaluations on 100,000 classes

In this section, we conduct experiments in the case where the number of categories reached 100,000.

*Corresponding author.

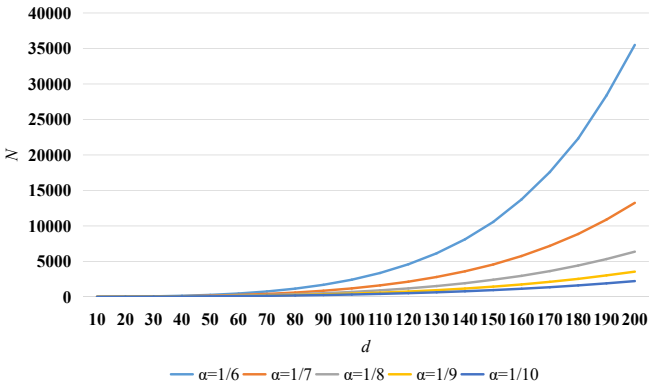

Figure 1. Relations between α, d and N.

Table 1. Experiments on the dataset with 100,000 classes. "Params." denotes parameters need to be optimized. "Top-1 Acc" represents Top-1 accuracy.

Method	Optimizer	EBVs Dim.	Params. (M)	Top-1 Acc (%)
FC	SGD	–	228.4	1.29
FC	AdamW	–	228.4	**30.25**
EBVs	SGD	5000	**33.8**	29.99

Dataset and settings We collect images containing 100,000 categories with almost the same number of training images as the ImageNet-1K dataset [2]. Specifically, we construct a dataset with 100,000 categories, each category contains 12 training images and 6 test images, *i.e.*, a total of 1.2 million images in the training set and 600,000 images in the test set. All these images and labels are collected from the citizen science website iNaturalist[1]. We adopt ResNet-50 as the backbone and follow Setting A1 in the paper. The hyper-parameters τ is set as 0.007 for our EBVs. All the models are pretrained on the ImageNet-1K dataset.

Results According to Table 1, we can see that ResNet-50 ending with a 100,000-way fully connected layer could not work when optimized with SGD [6]. The top-1 accuracy

[1] www.inaturalist.org

Table 2. Top-1 accuracy of ResNet-32 on the long-tailed CIFAR-10 and CIFAR-100 datasets.

Dataset	Long-tailed CIFAR-10			Long-tailed CIFAR-100		
Imbalance ratio	100	50	10	100	50	10
FC	38.32	43.85	55.71	70.36	74.81	86.39
EBVs	**40.41**	**44.68**	**57.82**	**73.31**	**78.97**	**87.9**

Table 3. Comparisons of classification accuracy (%) on the iNaturalist 2018 dataset.

Method	Test size	Top-1 Acc (%)
FC*	224^2	61.7
FC	224^2/256^2	64.03 / 65.43
EBVs	224^2/256^2	**65.00 / 67.12**

* denotes the model is trained without TrivialAugment and LR optimizations.

is only 1.29% after training for 105 epochs. When training with the AdamW [5] optimizer, the top-1 accuracy turns out to 30.25%. However, the 100,000-way fully connected layer contains around 200M parameters which is too large and will become huger if the number of categories continues to grow. When training with our proposed EBVs, if the dimension of each basis vector is set as 5,000, the final top-1 accuarcy gains 29.99%, while the parameters to be optimized are only 33.8M, which are only around $\frac{1}{7}$ parameters of previous models.

3. Empirical evaluations on long-tailed image classification

3.1. Datasets and settings

Long-tailed CIFAR-10 & CIFAR-100 Both CIFAR-10 and CIFAR-100 has 60,000 images of size 32×32 with 50,000 for training and 10,000 for validation. We choose the long-tailed version of CIFAR-10 and CIFAR-100 [1], which downsamples the training data class-wisely from the original dataset by exponential decay functions. For fair comparisons, imbalance factors we use in experiments are 10, 50 and 100.

iNaturalist 2018 iNaturalist 2018 [7] is a large-scale real-world dataset with 437,513 images from 8,142 categories. It naturally follows a severe long-tailed distribution with an imbalance factor of 512. Besides the extreme imbalance, it also faces the fine-grained problem [9]. In this paper, the official splits of training and validation images are utilized for fair comparisons. We utilize ResNet-50 [4] as the backbone.

Settings For long-tailed CIFAR-10 and CIFAR-100 datasets, we follow the data augmentation strategies proposed in [4]: randomly crop a 32×32 patch from the original image or its horizontal flip with 4 pixels padded on each side. we use ResNet-32 [4] as the backbone. SGD optimizer with momentum of 0.9 and weight decay of 5×10^{-4} is used for network optimization. We train all the models for 200 epochs with batch size of 128. For the iNaturalist 2018 dataset, we utilize ResNet-50 [4] as the backbone, the hyper-parameters τ is set as 0.02. We train the model by following Setting A1 in the paper, the training epoch is set as 200. The dimension of our proposed EBVs is set as 10, 100 and 8,142 for CIFAR-10, CIFAR-100 and iNaturalist 2018, respectively.

3.2. Results

We conduct extensive experiments on long-tailed CIFAR datasets with three different imbalanced ratios: 10, 50 and 100. Table 2 reports the top-1 accuracy of models ending with a general k-way fully cinnected layer and our proposed EBVs. EBVs outperform the general FC baseline in all the settings. In Table 3, we report the top-1 accuracy on the iNaturalist 2018 dataset. EBVs also gain around 1% improvement in all the settings.

Table 4. Ablation studies of the performance of stacked incremental improvements on top of baseline of our proposed EBVs. w/o EBVs denote models ending with a general fully connected classifier. ResNet-50 baseline is under Setting A0 in the paper but with only 1-crop testing. "Top-1 Acc" denotes Top-1 accuracy while "Abs. Diff." denotes absolute difference. The test size for each image is set as 224^2 except "FixRes Mitigations".

	Top-1 Acc (%)	Abs. Diff.
ResNet-50 Baseline	76.13	0.00
+ LR Optimizations w/o EBVs	76.49	0.36
+ TrivialAugment w/o EBVs	76.81	0.68
+ TrivialAugment	77.26	1.13
+ Random Erasing	77.55	1.42
+ Label Smoothing	77.61	1.48
+ Mixup	77.79	1.66
+ Cutmix	78.14	2.01
+ Long Training w/o EBVs	79.51	3.38
+ Long Training	79.73	3.60
+ FixRes Mitigations	**80.90**	**4.77**

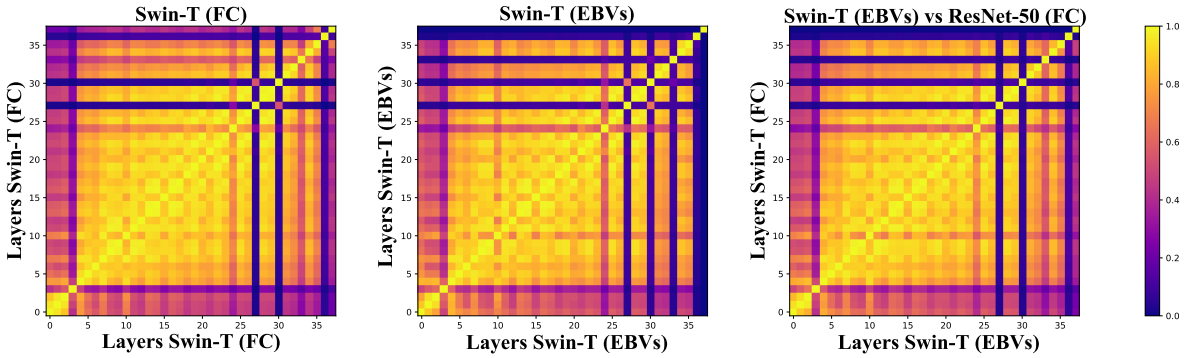

Figure 2. Representation structure of Swin-T. **Left:** Similarity between layers within Swin-T ending with fully connected layer with softmax. **Middle:** Similarity between layers within Swin-T ending with EBVs. Only last few layers share minimal similarity with other layers. **Right:** Similarity between layers across Swin-T ending with general fully connected layer with softmax and our proposed EBVs. Only last few layers share minimal similarity with other layers.

4. Ablation studies on training techniques

In this section, we conduct ablation studies of the performance of different training techniques in our proposed EBVs. As training models are not a journey of monotonically increasing accuracies and the process involves a lot of backtracking [8]. To quantify the effect of each optimization in our proposed EBVs, we conduct related ablation studies in Table 4. When the training crop size is fixed as 224^2 and turns the inference resolution to 320^2, with only 1-crop testing, EBVs gains a final top-1 accuracy of 80.9% on the ImageNet-1K dataset.

5. Do EBVs perform like FC?

In this section, we follow Section 5 of the paper and pick Swin-T as the backbone. As shown in Figure 2 in the supplementary materials, when adopting Swin-T as the backbone, the phenomenon of models ending with EBVs in the last few layers is similar to the performance in ResNet-50. However, most of the other layers share high similarities whether the model ends with a fully connected layer or EBVs.

References

[1] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. In *Adv. Neural Inform. Process. Syst.*, pages 1567–1578, 2019. 2

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 248–255, 2009. 1

[3] Michael Elad. *Sparse and redundant representations: from theory to applications in signal and image processing*, volume 2. Springer, 2010. 1

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 770–778, 2016. 2

[5] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 2

[6] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, pages 400–407, 1951. 1

[7] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The iNaturalist species classification and detection dataset. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8769–8778, 2018. 2

[8] Vasilis Vryniotis. How to train State-of-The-Art models using TorchVision's latest primitives. https://pytorch.org/blog/how-to-train-state-of-the-art-models-using-torchvision-latest-primitives/, 2021. 3

[9] Xiu-Shen Wei, Yi-Zhe Song, Oisin Mac Aodha, Jianxin Wu, Yuxin Peng, Jinhui Tang, Jian Yang, and Serge Belongie. Fine-grained image analysis with deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(12):8927–8948, 2022. 2