

A APPENDIX A

A.1 PLANCKIAN JITTER

Figure 4 illustrates the illuminants sampled from the distribution of a black body radiator, with correlated color temperature T in the interval between $3000K$ and $15000K$. The resulting spectra are visualized on the left and in the middle, while the resulting distribution of illuminants is visualized in the Angle-Retaining Chromaticity diagram on the right.

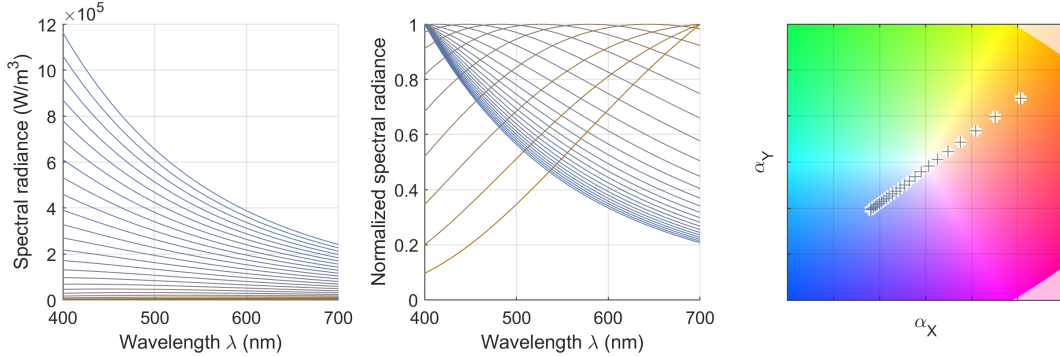


Figure 4: Spectral power distributions (left) and corresponding ARC chromaticities (right) of the sampled black body radiator, used to generate Planckian jittering.

Figure 5 shows a comparison between default color jitter (left) and Planckian jitter (right), replicating Figure 1 in xy chromaticity.

A.2 DATASET DETAILS

In section 4.4 of the main paper we analyzed the impact of our data augmentation when using the features extracted from the backbone trained on IMAGENET on new datasets. The datasets used in the finetuning step are:

- FLOWERS-102 (Nilsback & Zisserman, 2008): Dataset consisting of 102 flower categories commonly occurring in the United Kingdom. Each class consists of between 40 and 258 images, for a total 8,189 images.
- VEGFRU (Saihui Hou & Wang, 2017): Dataset consisting of more than 160,000 images of vegetables and fruits divided in 292 classes.
- CUB-200 (Welinder et al., 2010): Dataset made of 6,033 images of 200 bird species.
- T1K+ (Cusano et al., 2021): Dataset of textures divided into 1129 classes and organized in 5 groups of 266 super classes. We adopted the 266 class labeling to finetune and test our models.
- USED (Ahmad et al., 2016): Dataset consisting of 14 categories of social events from around the world. Images depict the interaction between multiple objects and the background scene. We considered 1000 images per class for training, and 500 images per class for testing.

A few example images for each of the color task datasets are given in Figure 6.

Additionally, in section A.6 of this appendix TINY-IMAGENET (Le & Yang, 2015) is used. It contains 100,000 images of 200 classes (500 for each class) at 64×64 pixel resolution.

A.3 COLOR SELECTIVITY INDEX

Color selectivity is defined by Rafegas & Vanrell (2018) as the property of a neuron that activates strongly when a specific color appears in the input image, and does not when the color is absent. It

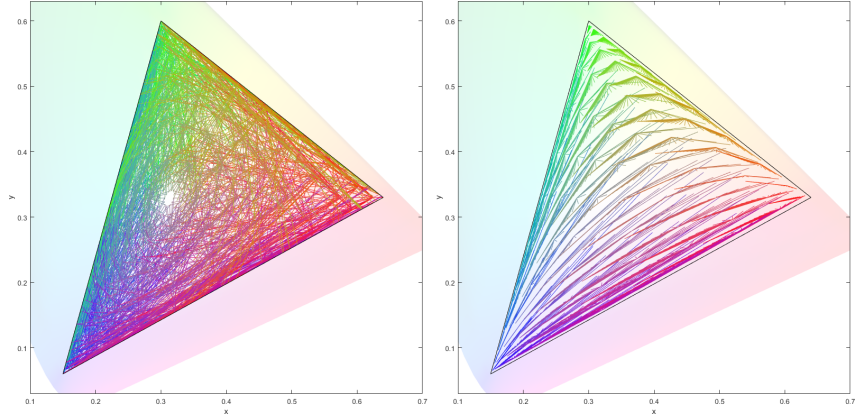


Figure 5: Default color jitter (left) and Planckian jitter (right) in xy chromaticity.



Figure 6: Example images from the datasets used as downstream classification tasks. From left to right: FLOWERS-102, CUB-200, VEGFRU, and T1K+.

is computed by estimating the ratio between the neuron’s global activation with color input images and the global activation with corresponding grayscale images:

$$\alpha(n^{L,i}) = 1 - \frac{\sum_{j=1}^N w'_{j,i,L}}{\sum_{j=1}^N w_{j,i,L}}. \quad (6)$$

Here $w_{j,i,L}$ refers to the activation of an image patch j for the i -th neuron $n^{L,i}$ at layer L , normalized for the maximum activation value across all possible image patches. $w'_{j,i,L}$ is the equivalent formulation for a grayscale version of the images. The set of considered image patches is restricted to the top- N regions from a given dataset that maximally activate the neuron of interest.

We can distinguish between neurons that are colorblind or neurons that highly rely on color information by looking at the α value obtained: an α value more than 0.25 means that the neuron is high color selective, while an alpha value less than 0.1 means that the neuron is basically colorblind. These thresholds were selected based on the analysis by Rafegas & Vanrell (2018). We collected alpha values for the neurons in the last layer of the encoders trained with different data augmentation configurations in order to compare the models sensitivity to color and how it changes in relation to the training procedure adopted.

A.4 COLOR SENSITIVITY

To analyze feature robustness to different illuminants, we tested the models with different, re-illuminated versions of the CIFAR-100 and FLOWERS-102 datasets. We applied *Planckian Jitter* on the two datasets, generating 25 different versions of each, one for each illuminant sampled. Using these different versions of the datasets we then test the models for each illuminant and collect the classification accuracies. The results on both CIFAR-100 and FLOWERS-102 are given in Figure 7.

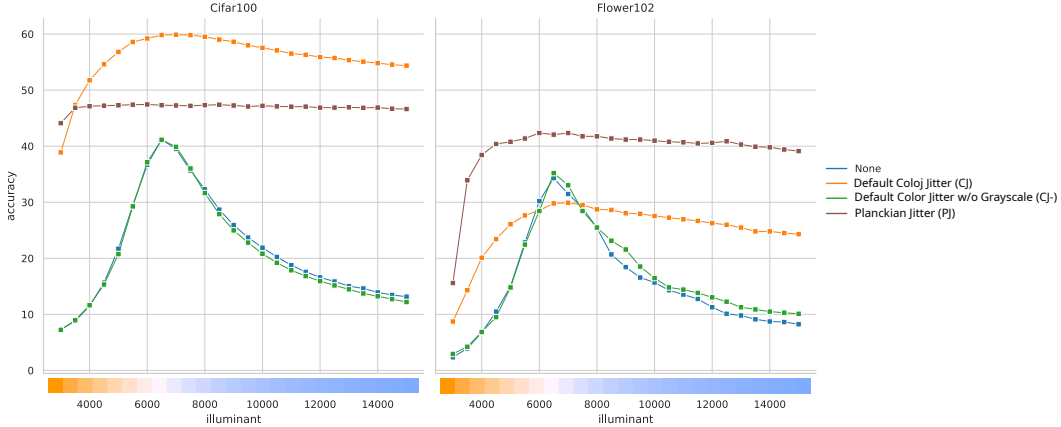


Figure 7: Illuminant robustness analysis. To assess feature invariance to realistic color changes in images, for each method we evaluate classification accuracy on 25 different, re-illuminated versions of the datasets. The images of the two datasets (CIFAR-100 on the left and FLOWERS-102 on the right) have been modified with the illuminants from temperature 3000 K to 15000 K using the Planckian Jitter transform.

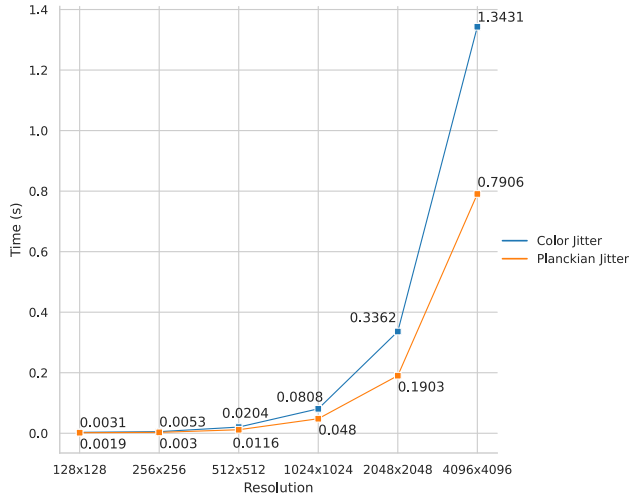


Figure 8: Comparison of execution time between the proposed *Planckian Jitter* transform and the Color Jitter implementation in Pytorch Torchvision v0.9.1. For each resolution we executed both the algorithms 40 times.

A.5 EXECUTION TIME COMPARISON

Here we provide an analysis of execution time to assess the usability of the Planckian Jitter compared to standard Color Jitter. We executed the two algorithms: the Color Jitter image transform from Torchvision (Torch version v1.8.1 and Torchvision version v0.9.1) and our *Planckian Jitter* at different image resolutions. For each resolution we ran the code 40 times and averaged the execution time. Results are shown in Figure 8. All augmentations were performed in CPU on an Intel i7-8700 processor. As can be seen, the proposed *Planckian Jitter* is faster than standard Color Jitter.

A.6 ADDITIONAL DOWNSTREAM RESULTS ON TINY-IMAGENET

We also performed experiments for several other configurations of the downstream tasks with the representation trained on Tiny-ImageNet. In Table 5 we report results for the main task and downstream task (as in section 4.4 of the main paper ImageNet, but here all images are at 64×64 pixel resolution).

Table 5: Additional analysis on downstream tasks. Self-supervised training is performed on TINY-IMAGENET at (64×64) .

DATA AUGMENTATION	TINY-IMAGENET	FLOWERS-102	CUB200	VEGFru	T1K+
None	27.06%	37.65%	18.76%	24.07%	35.82%
Default Color Jitter (CJ)	33.12%	46.27%	19.36%	23.92%	26.01%
Default Color Jitter w/o Grayscale (CJ-)	31.62%	40.39%	21.90%	27.39%	32.50%
Planckian Jitter (PJ)	30.95%	52.35%	25.12%	28.94%	32.51%
LSC: [CJ,CJ-]	39.02%	58.33%	26.82%	36.43%	37.20%
LSC: [CJ,PJ]	39.23%	61.57%	30.45%	39.65%	38.20%

Table 6: Classification accuracy as a function of down-sampling size on Flowers. Results confirm that PJ is less sensitive to down-sampling than CJ.

Method	32	64	128	224
CJ	7.74	49.23	91.43	91.90
CJ+PJ	16.30	66.43	93.01	93.45
PJ	23.11	70.13	89.04	89.22

These additional comparisons confirm the conclusions described in section 4.4 of the main paper. For all of the considered downstream tasks the application of the proposed data augmentation procedure improves the results even in comparison with other combinations of the originally used data augmentations. Moreover, the comparison with the latent space combination with the two versions of the default color jitter shows how exploiting features extracted by the model trained using the proposed Planckian Jitter augmentation enriches the expressive power of the final model.

A.7 DEPENDENCE ON RESOLUTION

To better understand the difference of the performances reported in Tables 2 and 3, we perform an experiment that shows that the relative importance of texture and shape increases with increased resolution.

As an additional experiment, we took the representations and classifiers learned at high-resolution (224×224) and investigated their sensitivity to high-frequency information in images. At inference time, we down-sample (down-sample resolution is given in Table 6) and then up-sample all images. In this way, we can compare the dependence on high-resolution information of different methods. Note, that here we do not retrain the classifier but use the one trained at 224×224 . The results clearly show that CJ suffers more from down-sampling than PJ. For the Flowers datasets, the results of CJ at resolution 224 are better than PJ. However, when we down-sample to 64, the results change and results for PJ are already significantly better than CJ. This suggest that the texture information (important for CJ) is removed and this hurts performance. For PJ, which is more dependent on color information, down-sampling hurts results less (note PJ is also using texture, shape but to a lesser degree, so results still deteriorate for smaller resolutions).

A.8 COLOR IMPORTANCE IN PLANCKIAN JITTER BASED REPRESENTATION

To verify that CJ uses less color information than PJ, we did a simple experiment where at inference time we changed the input images from sRGB to gray-scale images. These results are provided in Table 7. These results clearly show that PJ is much more dependent on color than CJ. PJ has a drop of over 67.6% whereas CJ only drops 3.2 percentage points.

Table 7: Methods evaluated with color and grayscale images on the Flowers dataset.

Method	COLOR	ACCURACY
CJ	COLOR	92.73
CJ	GS	89.51
PJ	COLOR	88.97
PJ	GS	21.38