WarpedGANSpace: Finding non-linear RBF paths in GAN latent space – Supplementary Material –

Christos Tzelepis, Georgios Tzimiropoulos, Ioannis Patras
Queen Mary University of London
Mile End road, E1 4NS London, UK
{c.tzelepis, g.tzimiropoulos, i.patras}@qmul.ac.uk

Ablation study

In this section, we will present our ablation study on the various components of the proposed method. More specifically, we will first conduct a study on the number of bipolar pairs of support vectors used for learning each interpretable non-linear path and will show that by increasing them we arrive at more accurate reconstructor $\mathfrak{R}$ (Fig. 2 in the manuscript) i.e., at paths that are discriminated more effectively and therefore are more likely to be more interpretable. Then, we will experimentally show that linear paths [1] can be derived as a special case of our method, as discussed in Sect. 3.2 in the manuscript. Finally, we will study the effect of the weighing parameter α (Eq. (1) in the manuscript); i.e., the weight of each support vector, and will show that optimizing the scale parameter γ, along with the positions of the support vectors for $\alpha_i = \{\pm 1\}$, leads to better results, and thus this is our choice for training our proposed method.

Number of bipolar support vector pairs

In Table 1 we show the accuracy of the reconstructor $\mathfrak{R}$ (Fig. 2 in the manuscript) for different numbers of bipolar pairs of support vectors and different choices of the weighting and scaling parameters α and γ, respectively. In general, we note that increasing the number of support vectors leads to better classification accuracy. This is an indication that the paths that are induced by more support vectors can be discriminated more effectively and therefore are more likely to be more interpretable. The table also shows that keeping the alphas fixed to $\alpha \in \{\pm 1\}$ leads to better accuracy, see also subsection *Weighting parameters* α.

Linear directions as a special case

As discussed in the manuscript (Sect. 3.2), linear interpretable paths (such as those obtained by [1]) can be derived using our method by learning one bipolar pair of support vectors per path, i.e., $N = 2$, with fixed and small γ scale parameters (in this case we set $\gamma = 10^{-6}$) and fixed weights $\alpha \in \{\pm 1\}$. This is clear in the Figure 1, where we show that all of the 200 paths are practically linear (black line). We may directly compare this with the case where we also learn one pair of bipolar support vectors ($N = 2$) with fixed $\alpha \in \{\pm 1\}$ and optimized γ parameters, which, as shown by the blue line in Fig. 1, leads to non-linear paths.

Table 1: Reconstructor accuracy of the proposed method for different choices of the number of support vector bipolar pairs (N), weighting parameter α (learned or fixed to ± 1), and scale parameter γ.

N	Learn α	Learn γ	Accuracy (%)
2 (1 bipolar)	✗	✗	87.1
2 (1 bipolar)	✗	✓	95.3
8 (4 bipolars)	✗	✓	95.1
8 (4 bipolars)	✓	✓	89.9
32 (16 bipolars)	✗	✓	98.8
32 (16 bipolars)	✓	✓	96.8
64 (32 bipolars)	✗	✓	**99.3**
64 (32 bipolars)	✓	✓	97.8

Weighting parameters α

In this section we will discuss how the weighing parameters α affect the obtained non-linear interpretable paths. First, in terms of reconstructor's classification accuracy, we note that learning both α and γ leads to slightly less accurate reconstruction compared to the case where only the scale parameters γ are optimized and α are set to $\{\pm 1\}$.

In Fig. 4 we show a histogram of the scale parameters γ in the cases where α's are optimized or set to $\{\pm 1\}$ during training, and in Fig. 2 the corresponding histograms that shows the non-linearity of the resulting paths – both show that in the latter case we arrive to paths that are more non-linear. In Table 1 we note that the accuracy of the reconstructor, which is our optimization criterion set as a proxy for the interpretability of the paths, is also higher in the latter case. This is an indication that non-linearity is impor-

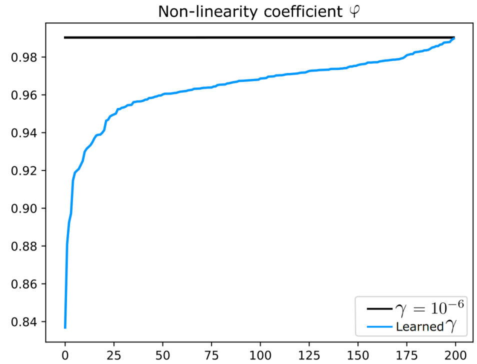

Figure 1: Illustration of the non-linearity coefficient φ for the discovered paths obtained by bipolar pairs of support vectors with fixed $\gamma = 10^{-6}$ (black line) compared to the corresponding paths induced by bipolar pairs of support vectors with learned γ parameters.

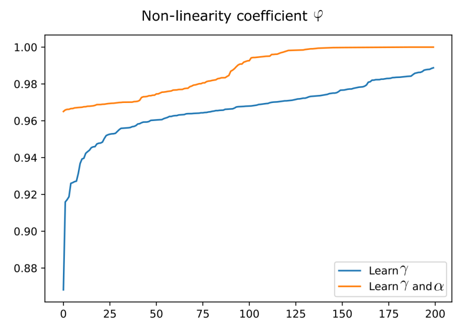

Figure 2: Comparison between non-linear paths induced by $N = 64$ support vectors (32 bipolars) each, in the cases where the scale parameter γ is optimized jointly with the weighting parameter α or not, in terms of the non-linearity coefficient φ.

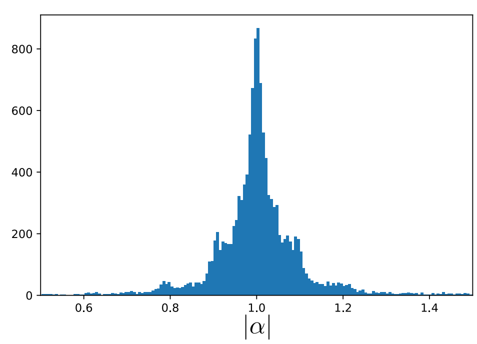

Figure 3: Distribution of weighting parameters α in case where they are jointly optimized with scale parameters γ for sets of $N = 32$ pairs of bipolar support vectors.

tant for better discrimination and this is the setting that we adopted for our experiments.

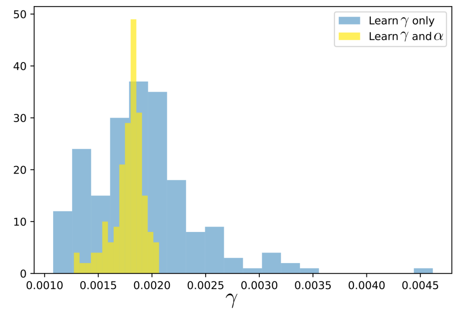

Figure 4: Distribution of scale parameters γ in the cases where a) they are jointly optimized with weighting parameters α (yellow bars) and b) they are optimized using fixed weighting parameters $\alpha \in \{\pm 1\}$.

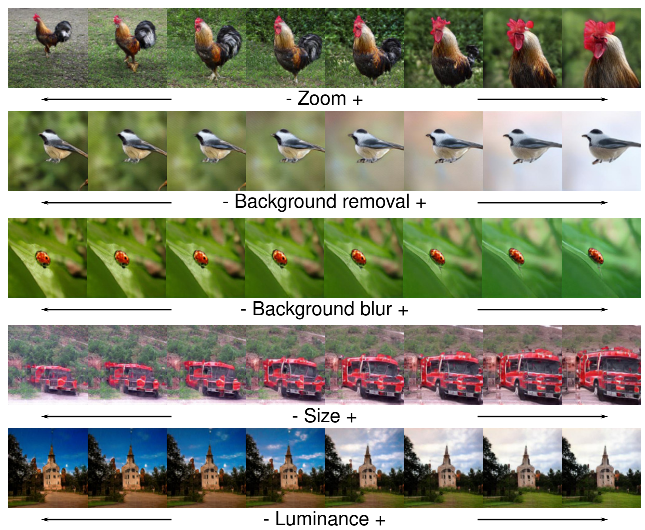

Figure 5: Examples of non-linear interpretable paths discovered by the proposed method on general images (BigGAN/ImageNet).

Results on general images datasets

In this section, we will briefly present additional results of the proposed method on general image datasets. More specifically, similarly to previous work [34], our method can work with other datasets/objects. In Fig. 5, we show some of the interpretable paths that the proposed method discovers (on various BigGAN/ImageNet classes).

References

[1] A. Voynov and A. Babenko. Unsupervised discovery of interpretable directions in the GAN latent space. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 9786–9796. PMLR, 2020. 1