Tensor Component Analysis for Interpreting the Latent Space of GANs

Supplementary Material

1 Introduction

In this document, we present additional material to support the main paper. Firstly, we provide illustrations and derivations in Section 2, aimed at clarifying and providing intuition into some of the operations performed in the main paper. Lastly, in Section 3 we provide experimental results designed to supplement and further validate our proposed method.

2 Illustrations & intuition

2.1 Mode-n edits

In the main paper, we form an 'edit tensor' $\mathcal{Z}' \in \mathbb{R}^{C\times H\times W}$ which is a combination of the basis vectors for each of the three modes of the generator's activations. We show how one can make edits that, broadly speaking, correspond to style or geometry by adding the mode-n basis vectors to all mode-n fibers of this edit tensor, using the 1^{st} order terms $\mathcal{Z}' = \mathcal{S}_n \times_n \mathbf{U}^{(n)}$.

To see how these 1^{st} order terms work to select the desired linear combinations of the N basis vectors from the columns of $\mathbf{U}^{(n)}$ and sum them along each of the output's mode-n fibers, we can inspect $\mathcal{Z}'$'s mode-n unfolding. We know from the definition of the mode-n (matrix) product [4] that we can write this term equivalently as

$$\mathcal{Z}' = \mathcal{S}_n \times_n \mathbf{U}^{(n)} \quad \Leftrightarrow \quad \mathbf{Z}'_{(n)} = \mathbf{U}^{(n)}\mathbf{S}_{n\,(n)}. \tag{1}$$

Next, recall that the definition of the mode-n unfolding of a tensor $\mathcal{X}$ is a rearranging of its mode-n fibers into the columns of a matrix $\mathbf{X}_{(n)}$ [4]. With this in mind, we can inspect the right-hand-side of Eq. (1), writing it as

$$\begin{aligned}
\mathbf{Z}'_{(n)} &= \mathbf{U}^{(n)}\mathbf{S}_{n\,(n)} && (2)\\
&= \underbrace{\begin{bmatrix} | & & | \\ \mathbf{u}_1^{(n)} & \cdots & \mathbf{u}_N^{(n)} \\ | & & | \end{bmatrix}}_{\mathbf{U}^{(n)}} \underbrace{\begin{bmatrix} - & \alpha_1 \mathbb{1}^\top & - \\ & \vdots & \\ - & \alpha_N \mathbb{1}^\top & - \end{bmatrix}}_{\mathbf{S}_{n\,(n)}} && (3)\\
&= \sum_i \mathbf{u}_i^{(n)} \circ \alpha_i \mathbb{1}, && (4)
\end{aligned}$$

which shows that each of the mode-n fibers of $\mathcal{Z}'$ are linear combinations of the mode-n basis vectors, as intended.

2.2 Multilinear mixing

We also show we can model the interactions of the basis vectors between the modes of the tensor. We first recall the following useful result with the Kronecker product [3]:

Proposition 1. *Let* $\mathcal{Y} = \mathcal{X} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \cdots \times_N \mathbf{U}^{(N)}$, *then*

$$\mathrm{vec}\left(\mathcal{Y}\right) = \left(\mathbf{U}^{(N)} \otimes \mathbf{U}^{(N-1)} \otimes \cdots \otimes \mathbf{U}^{(1)}\right) \mathrm{vec}\left(\mathcal{X}\right). \tag{5}$$

The 3$^{\text{rd}}$ order term $\mathcal{Z}' = \mathcal{S}_{CHW} \times_1 \mathbf{U}^{(C)} \times_2 \mathbf{U}^{(H)} \times_3 \mathbf{U}^{(W)}$ can then be understood most easily by appealing to Proposition 1 and writing it in terms of its vectorisation as

$$\begin{aligned} \mathrm{vec}\left(\mathcal{Z}'\right) &= \left(\mathbf{U}^{(W)} \otimes \mathbf{U}^{(H)} \otimes \mathbf{U}^{(C)}\right) \mathrm{vec}\left(\mathcal{S}_{CHW}\right) & (6) \\ &= \left(\mathbf{U}^{(W)} \otimes \mathbf{U}^{(H)} \otimes \mathbf{U}^{(C)}\right) \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{C \cdot H \cdot W} \end{bmatrix}}_{\mathrm{vec}(\mathcal{S}_{CHW})}. & (7) \end{aligned}$$

That is, considering the operation in its vectorised form, the 'selector tensor' $\mathcal{S}_{CHW}$ can be interpreted as simply taking a linear combination of the columns of the matrix formed from the interactions of the basis vectors of the three factor matrices. For example, $\mathrm{vec}\left(\mathcal{S}_{CHW}\right)(1) := \alpha_1$ weights the interactions of the first basis vectors of all three of the bases $\mathbf{u}_1^{(C)}, \mathbf{u}_1^{(H)}, \mathbf{u}_1^{(W)}$.

2.3 Regressing the edit tensor to the latent code

Finally, we illustrate how these terms are summed to form the edit tensor graphically in Fig. 1. The 'generalised inner product' then maps this back to the latent code $\mathbf{z}' \in \mathbb{R}^d$.

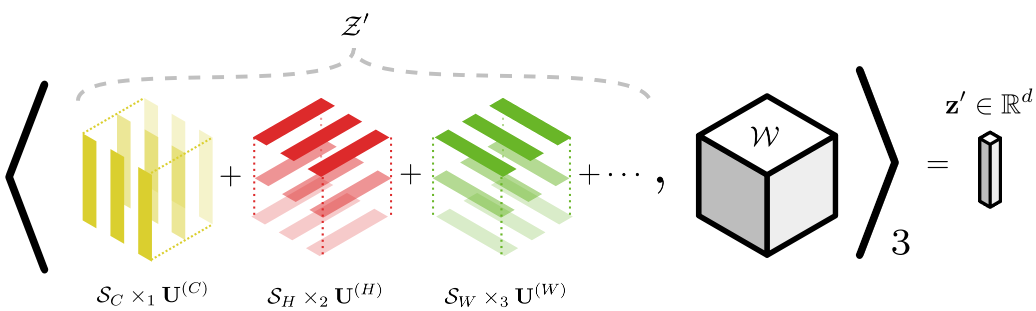

Figure 1: An overview of how we form the edit tensor and compute its latent code: We first form the edit tensor $\mathcal{Z}'$, the 1$^{\text{st}}$-order terms being demonstrated graphically here. Then, we take the generalised inner product with weight tensor $\mathcal{W}$ yielding $\mathbf{z}'$: the corresponding direction in the original latent space.

2.4 Computing the bases in a pre-trained GAN

To compute the three bases $\mathbf{U}^{(C)}, \mathbf{U}^{(H)}, \mathbf{U}^{(W)}$ we follow [1], where Lu et al. show that, if we retain each of the basis vectors (i.e., perform no dimensionality reduction), we can compute these factor matrices in one-shot. Given a pre-trained GAN's generator G, we compute the bases following Algorithm 1.

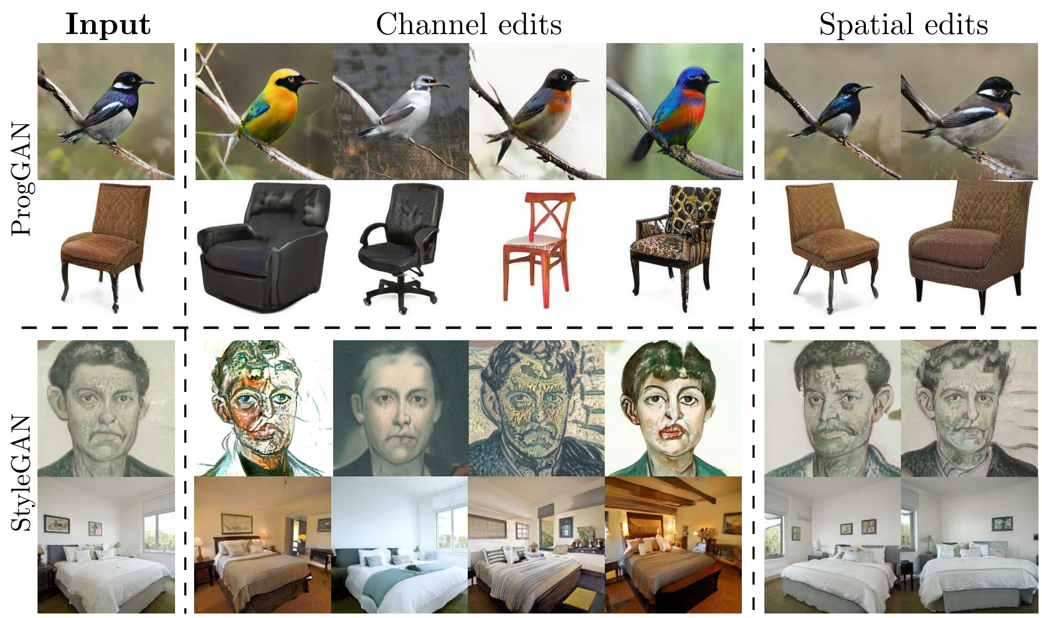

Figure 2: Edits performed along the spatial and channel modes separately, in a variety of generators and datasets. For these experiments, we use a low-rank Tucker decomposition for the regression tensor.

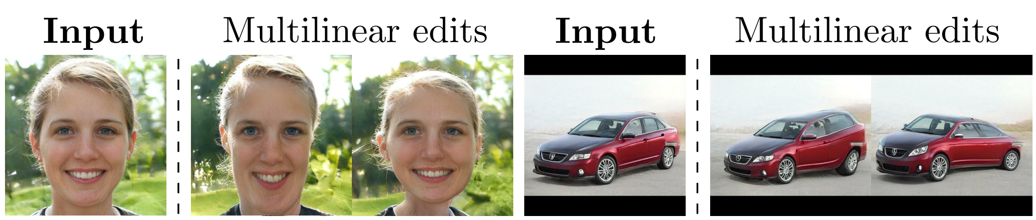

Figure 3: Edits found in the third-order interactions of bases for StyleGAN (when we directly edit the activation tensor).

Algorithm 1 Computing the (full rank) multilinear bases

procedure ComputeBases(G, i) ▷ Pretrained generator G and target layer i
  $\mathbf{Z} \leftarrow$ Sample M times from standard normal
  $\mathcal{Z} \leftarrow G[:i](\mathbf{Z}) \in \mathbb{R}^{M \times C \times H \times W}$ ▷ Intermediate activations
  for $n = 1:3$ **do**
    $\bar{\mathbf{Z}}_{(n)} \leftarrow \frac{1}{M} \sum_{m=1}^{M} \mathbf{Z}_{m(n)}$ ▷ Mean mode-n unfoldings
    $\mathbf{U}^{(n)} \leftarrow$ Left-singular vectors of $\sum_{m=1}^{M} \left(\mathbf{Z}_{m(n)} - \bar{\mathbf{Z}}_{(n)}\right) \left(\mathbf{Z}_{m(n)} - \bar{\mathbf{Z}}_{(n)}\right)^{\top}$
  end for
  return $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}$
end procedure

3 Additional experimental results

3.1 Qualitative results

Here we provide additional qualitative results. We first show more mode-wise edits in Fig. 2. In Fig. 3 we also perform 'multilinear mixing' on StyleGAN. For StyleGAN, we find these multilinear directions are best realised when we additionally add the edit tensor to the activation tensor directly–along with generating the corresponding latent code for the AdaIN operations–and continue the forward pass from this edited tensor.

3.2 Experimental setup

Here we describe more thoroughly the experimental setup we use to produce our results in the quantitative comparisons in the main paper, to ensure reproducibility. In general, we walk along each direction by a manual amount for each direction of each baseline, such that the total average mean change in predictions for each attribute is as close as possible. E.g. we walk along the 'blonde hair' direction for all methods until the mean change in this attribute is close to 0.5. The baselines we benchmark our proposed method against are detailed below.

GANSpace For GANSpace [2], no official weights are provided for the CelebA-HQ dataset for the ProgGAN generator. Therefore we manually implement this on top of the author's official code[1], using a total of $1,000,000$ samples to perform the decomposition.

SeFA For SeFA [5], we use the author's official code and pre-trained weights[2] to produce the edited images, manually identifying the directions that most closely correspond to the three attributes of interest. Concretely, we use indices `2,4,7` of the directions matrix for the attributes 'yaw', 'blond hair', and 'pitch' respectively.

Unsupervised discovery For [6], the official weights provided are trained on a different ProgGAN model to the other baselines. For this reason, we generate and use the predictions on a new ground-truth training set from this pre-trained model for this baseline for fair comparison. Using the pre-trained weights, we use indices `5,12,49` of the directions matrix for the attributes 'blond hair', 'yaw', and 'pitch' respectively.

Ours For our method, we use the indices `3,3,1` of the channel, height, and width bases for the attributes 'blond hair', 'pitch', and 'yaw' directions respectively.

3.3 Ablation study: choice of rank for $\mathcal{W}$

We briefly turn our focus to exploring the role of the regression tensor $\mathcal{W}$. We find the regularisation afforded by the choice of rank in the decomposition in regression tensor $\mathcal{W}$ to play an important role in the kind of edits we can generate. For example, we show in Fig. 4 the head-thinning multilinear direction using both a high- and low-rank Tucker structure on the regression tensor. We find a high rank necessary to generate the 'multilinear mixing' edits (that feature transformations very far from the true data distribution). However for smooth, interpretable directions such as pitch and yaw, we find a low-rank necessary–we find a high-rank can lead to artefacts for the first-order terms. Our findings here suggest that the rank of the decomposition of the tensor regression weight is a hyperparameter that one can tune depending on the types of edits they wish to generate. In practice we regress these terms back to the latent code separately.

Figure 4: An example of a multilinear direction affecting the head width, generated using both a high- (a) and low-rank (b) regression tensor.

[1] GANSpace codebase: https://github.com/harskish/ganspace

[2] Official SeFA weights: https://github.com/genforce/sefa/blob/master/latent_codes/pggan_celebahq1024_latents.npy

References

[1] Haiping Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. "MPCA: Multilinear Principal Component Analysis of Tensor Objects". In: *IEEE Trans. Neural Netw.* 19 (2008), pp. 18–39.

[2] Erik Härkönen et al. "GANSpace: Discovering Interpretable GAN Controls". In: *Proc. Adv. Neural Inform. Process. Syst. (NeurIPS)*. Vol. 33. 2020, pp. 9841–9850.

[3] Tamara G. Kolda. *Multilinear Operators for Higher-order Decompositions*. Tech. rep. Sandia National Laboratories, 2006.

[4] Tamara G. Kolda and Brett W. Bader. "Tensor Decompositions and Applications". en. In: *SIAM Review* 51.3 (Aug. 2009), pp. 455–500.

[5] Yujun Shen and Bolei Zhou. "Closed-Form Factorization of Latent Semantics in GANs". In: *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*. 2021.

[6] Andrey Voynov and Artem Babenko. "Unsupervised discovery of interpretable directions in the GAN latent space". In: *Proc. Int. Conf. Mach. Learn. (ICML)*. 2020, pp. 9786–9796.