# Do Computer Vision models internally differentiate visual and conceptual aspects of art without explicit supervision?

In the context of Deep Learning, disentanglement learning was developed with the aim of discerning controllable factors of variation in the latent representations of data models [1,2,3]. Identifying such marginal independences in the latent representations and enforcing learning in this direction favours the formation of meaningful latent representations, and the isolation of crucial abstractions in models. With the widespread adoption of generative AI models to create artistic images, it is important to identify the affordances of such models with respect to what abstractions they internally learn, what types and what explicit supervisions induce which learning.

We perform a supervised disentanglement experiment on StyleGAN3 [4], a state of the arts model for disentaglement learning, trained from scratch on the iMET textile collection dataset [5] at the task of image reconstruction. The experiment searches for independent factors of variations ranging from the colour and shape level features to the location and time-period of the textile in the image. For each feature, whenever possible, we find an approximation of the disentangled direction (a vector in the latent space of the model), using low-level image processing features for the first type, and training an image classifier on the existing metadata of the iMET dataset for the high level features. We experiment on textile data due to their prevalently bidimensional nature and the abundance of visual patterns it offers. Furthermore, in this work, we do not introduce any external supervision or guidance to the classifier, which will be object of future work. On the contrary, we investigate the disentangled features of a purely visual model, assessing which levels of understanding are independently encoded in the latent space of such model and can be manipulated.

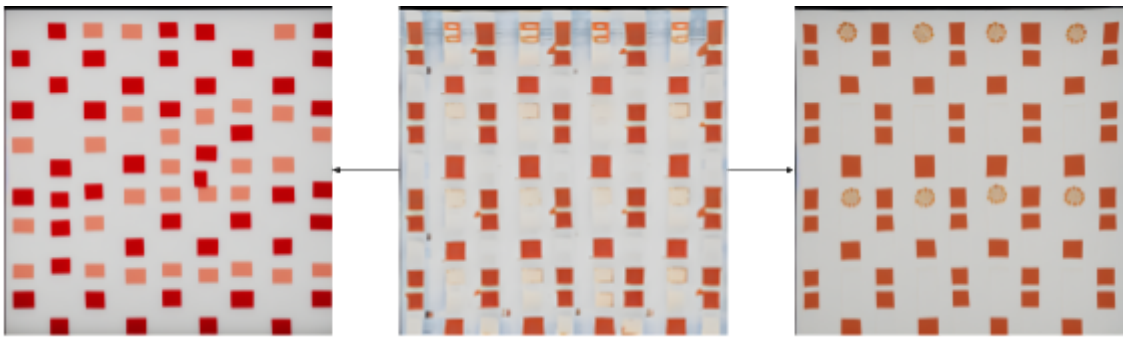

Image transformed moving in the disentangled direction of *pink* colour.

Original image generated by StyleGAN3 trained on iMET textile dataset.

Image transformed moving in the disentangled direction of *Art Nouveau*.

Figure 1: Example of disentangled features on textile images. On the left, a manipulation of the original image in the direction of a low level disentangled direction; on the right, the manipulation of a high-level disentangled feature.

## References


1. Do, Kien, and Truyen Tran. “Theory and Evaluation Metrics for Learning Disentangled Representations,” 2020.
2. Bengio, Joshua. From Deep Learning of Disentangled Representations to Higher-Level Cognition, 2018. https://www.youtube.com/watch?v=Yr1mOzC93xs.
3. Shen, Yujun, Ceyuan Yang, Xiaoou Tang, and Bolei Zhou. “InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs.” IEEE Transactions on Pattern Analysis and Machine Intelligence 44, no. 4 (April 2022): 2004–18. https://doi.org/10.1109/TPAMI.2020.3034267.
4. Karras, Tero, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. “Alias-Free Generative Adversarial Networks.” arXiv, October 18, 2021. https://doi.org/10.48550/arXiv.2106.12423.
5. “IMet Collection 2019 - FGVC6.” Accessed July 24, 2023. https://kaggle.com/competitions/imet-2019-fgvc6.