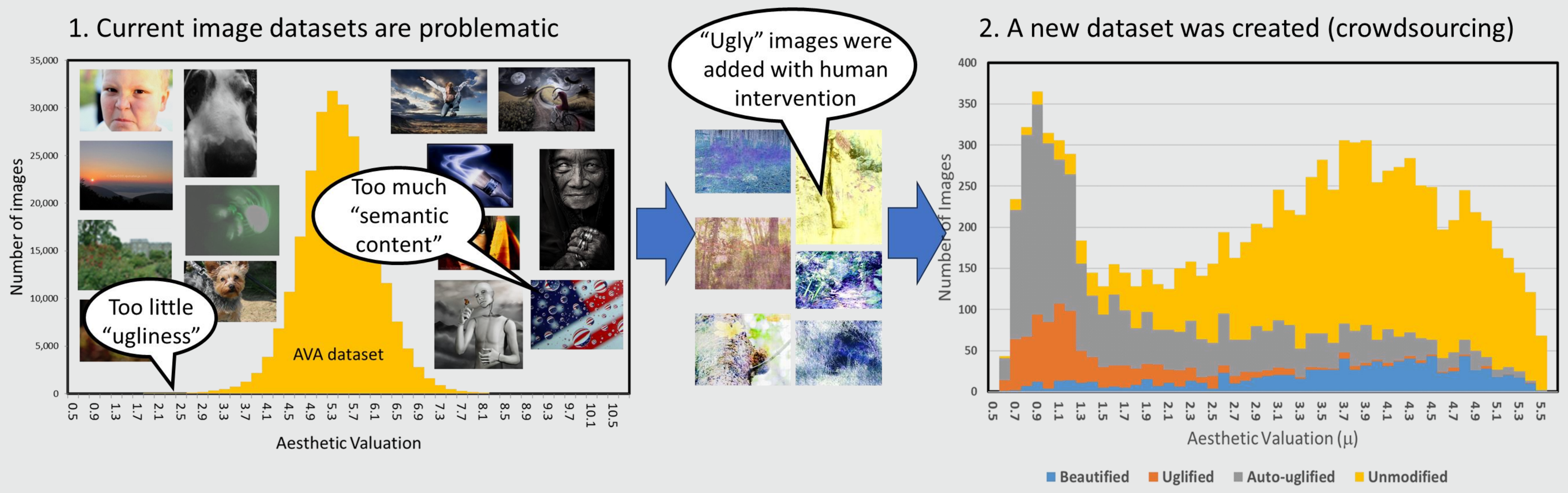
C. Alejandro Parraga*†, Arslan Javed*†, Bogdan Raducanu†
*Dept. Ciències de la Computació / †Computer Vision Center,
Universitat Autonoma de Barcelona (UAB), Spain.

Synthesizing Aesthetic Diversity: Creating a Balanced Dataset of Visual Aesthetics Using GANs

Motivation: Unbalanced and Semantic-rich image datasets

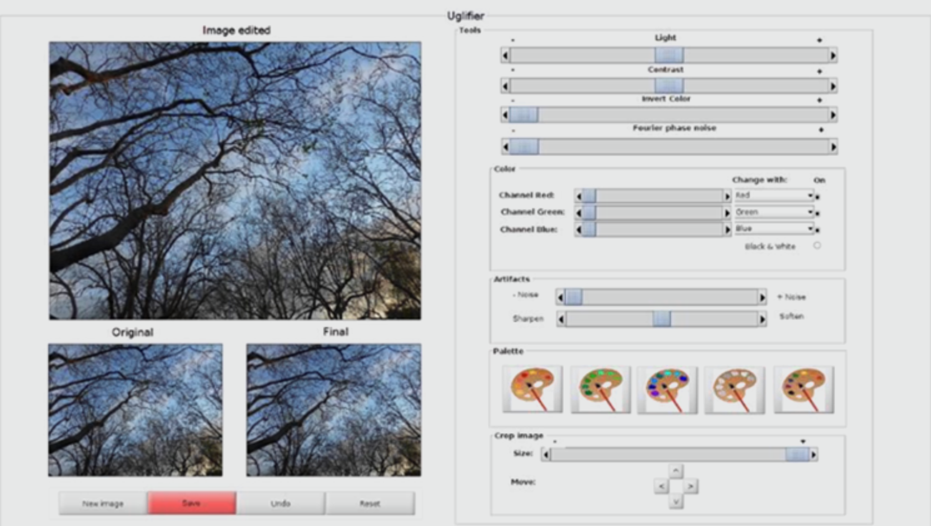
- **Addressing a Neglected Aspect:** The field of visual aesthetics primarily emphasizes the study of beautiful and visually appealing images, neglecting the aspect of ugliness in existing datasets.
- **Critical for Computer Vision Systems:** Neglecting the perception of ugliness poses challenges for robust computer vision systems. A limited dataset on ugliness restricts the capacity of different methodologies to interpret and distinguish images considered aesthetically ugly, hindering advancements in this domain.
- **Sophisticated Computer Vision Systems:** Developing methodologies for generating and analyzing a comprehensive dataset of ugly images contributes to the creation of more sophisticated and inclusive computer vision systems that can effectively handle both aesthetically pleasing and ugly images.



Methods

Training dataset:

The "uglifier" allows users to modify images using various techniques (chromatic changes, achromatic changes, edge filtering, Fourier statistics manipulation, etc.) These modified images contributed to reduce the bias towards beauty of the training dataset [1].



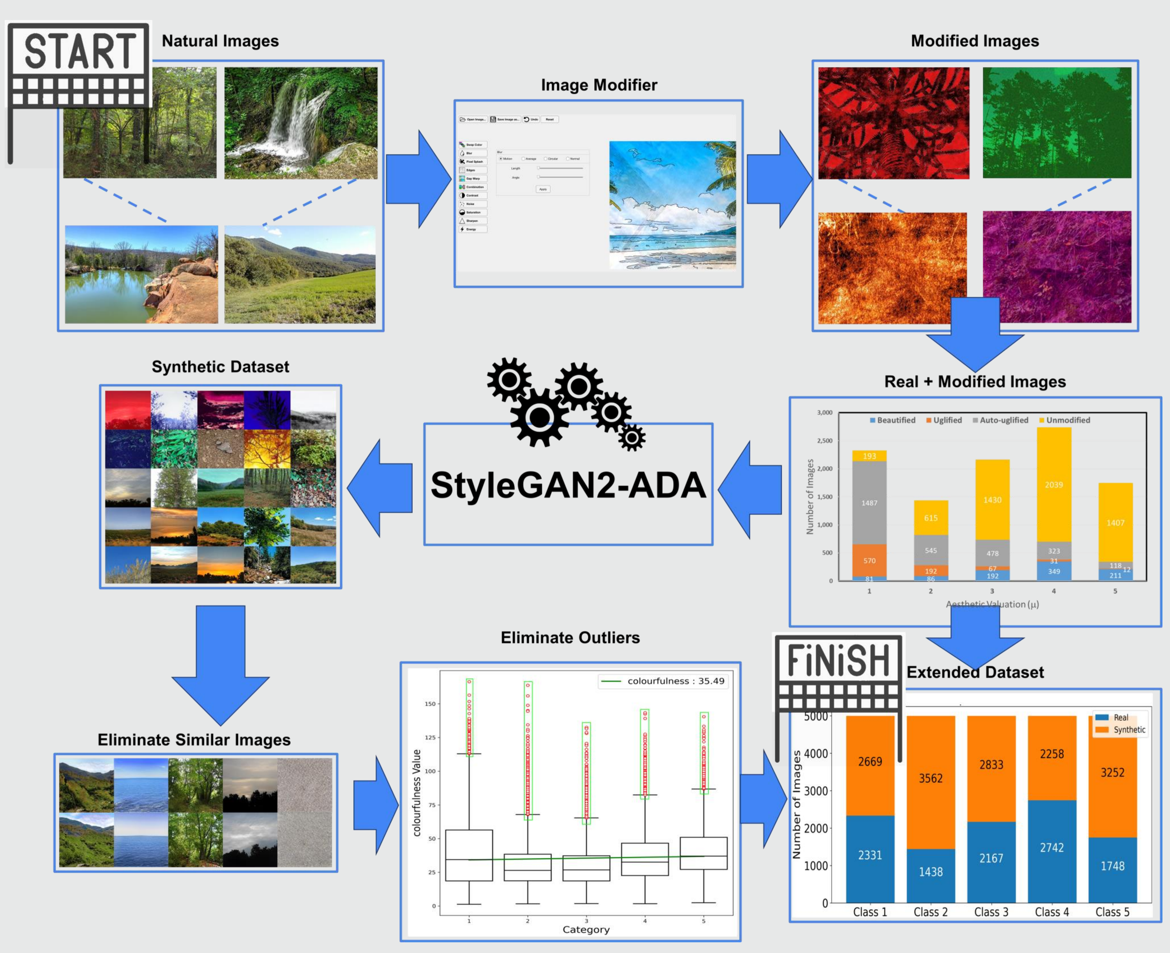
Collection and Categorization of Data:

- We collected 5684 natural images from different sources.
- We modified 4742 images using *The Uglifier*.
- Each Image was evaluated by 100 individuals to obtain the aesthetic score (crowdsourcing).
- Images were split into 5 classes according to their aesthetic value (1= very ugly, 5= very beautiful).

Dataset expansion:

The dataset was expanded by training a Generative Adversarial Network (GAN) [2] to produce images in each of the five aesthetic categories (1= very ugly, 5= very beautiful).

Workflow



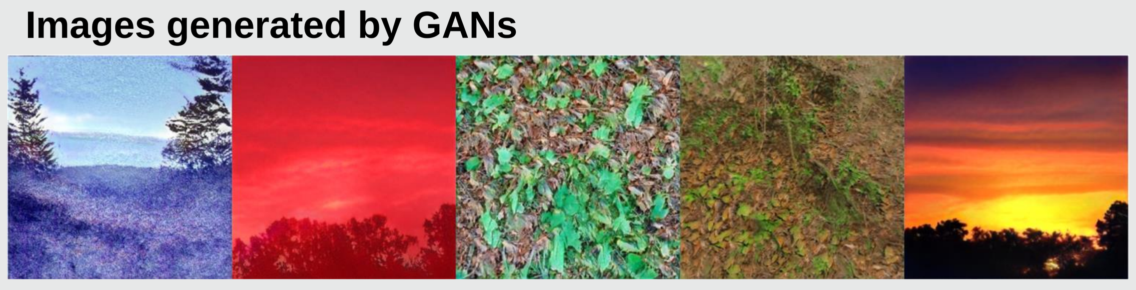
Use of StyleGAN2-ADA:

- First, we resize and crop the center of the images (512 x 512 pixels).
- The original dataset used for training StyleGAN2-ADA consists of 10426 images split into 5 classes from v.ugly to v.beautiful.
- Implemented different augmentations (Blit, Geom, Filter, Noise, Cutout) to get the best results.
- Initially, 8000 images were generated for each class.

Examples:

Images modified by humans

Images generated by GANs



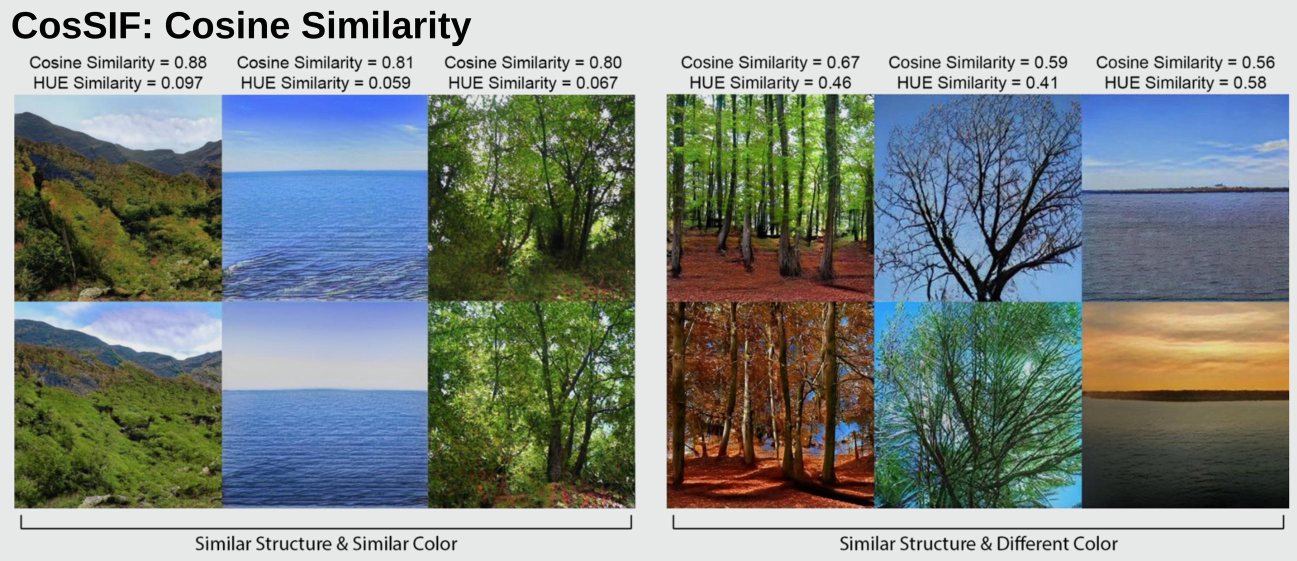
Similarity check for Real and Synthetic Datasets:

- Synthetic images that were too similar to another existing image were removed.
- Synthetic images with low-level properties too different from those expected for their class (outliers) were removed.
- The final dataset consisted of 5000 images in each class (counting both, synthetic and real)

Results

Similarities in the generated images:

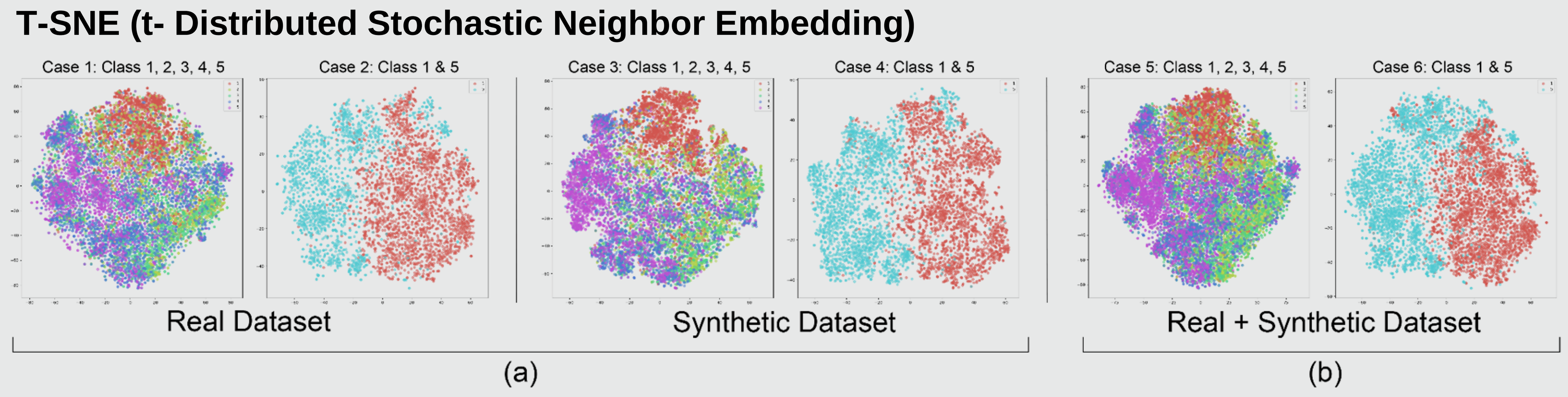
CosSIF: Cosine Similarity



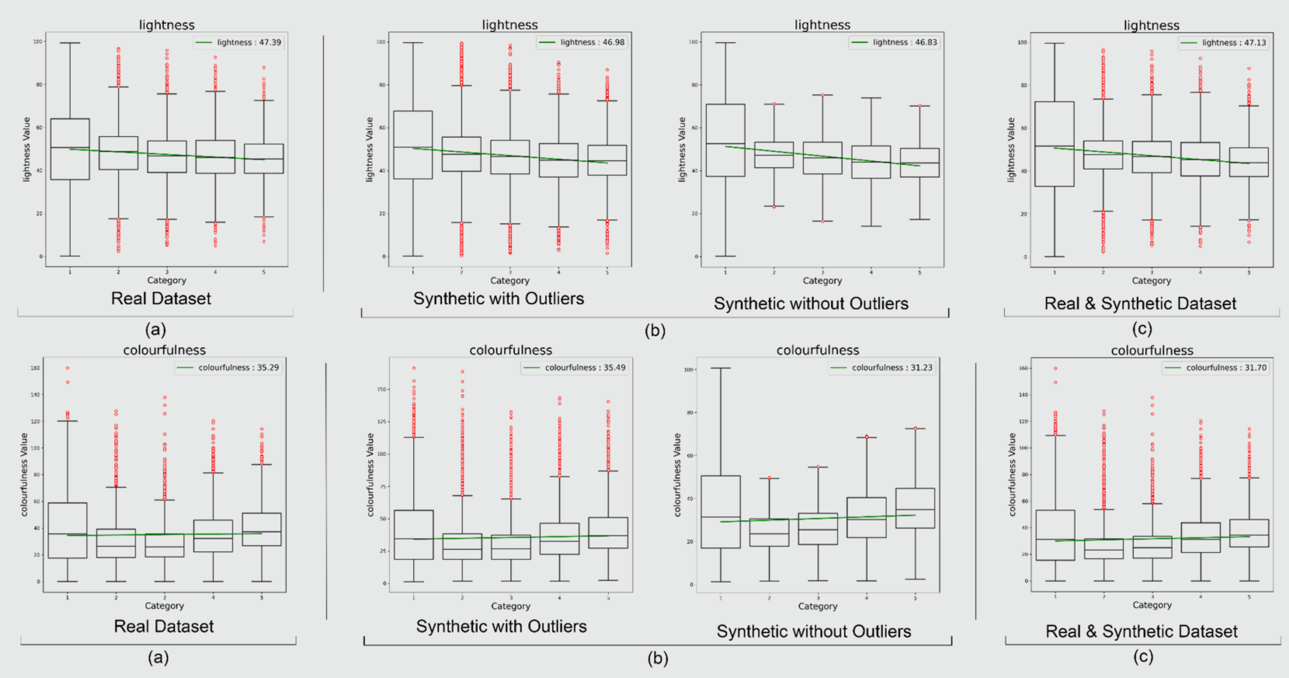
CosSIF: Cosine Similarity

- A technique for detecting and removing images that are structurally similar to other images
- Removes training images that are similar to images of other classes
- Removes images with less discriminative features after training

T-SNE (t- Distributed Stochastic Neighbor Embedding)



Boxplots of Real, Synthetic and Real + Synthetic Datasets



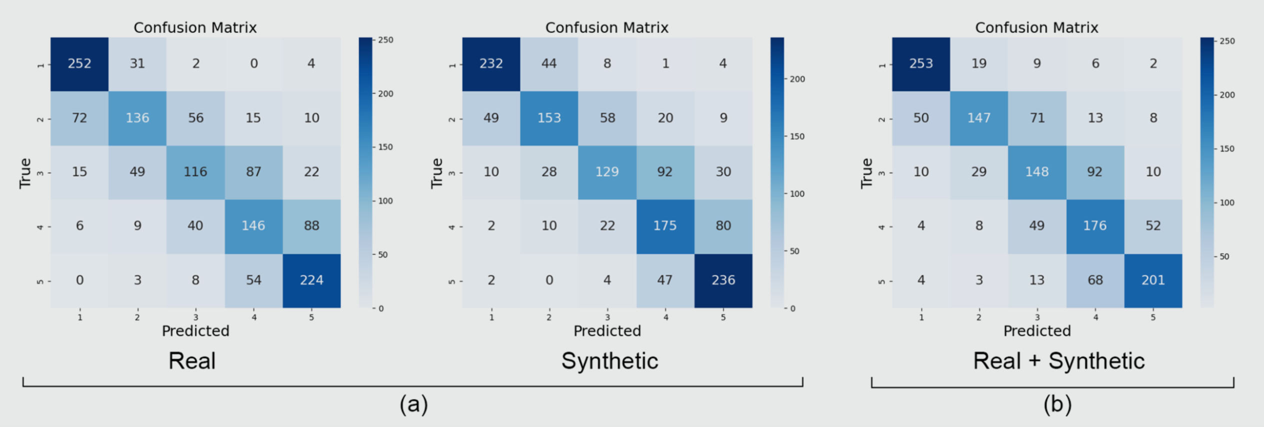
T-SNE: Similarity check of Real and Synthetic images

- T-SNE [4] is a data visualization technique to display high-dimensional data by giving each datapoint a location in a low-dimensional map.

Outliers removal:

- Synthetic images whose *low-level statistics* were too dissimilar to those of the real images in their class (ouliers) were also removed.
- Low-level statistics refer to Global Contrast, Colourfulness, Blur, Saturation, Lightness, Symmetry, Colour Gamut, Depth, etc. to ensure the same properties in both real and synthetic datasets

Confusion Matrices for a trained classifier (ResNet-18)
(Trained on ImageNet and I fine-tuned it on our Data)



Conclusions and Future Work

- By applying several similarity metrics, we can conclude that the synthetic images have similar properties as the real ones.
- We generated synthetic images to enlarge and balance each class by adding these images.
- By eliminating all the outliers we are able to get the best generated images for each class of our synthetic dataset.
- Future work: Comparative analysis of Ugly and Uncomfortable Images. In the next phase of our research, we plan to explore the properties of ugly and uncomfortable images to understand the distinctions and overlaps between these two aspects of visual perception. This comparison will involve a detailed analysis of the cognitive and emotional responses these images provoke.

Acknowledgements & References

- This research is part of the R+D+I grant PID2020-118254RB-I00 financed by MCIN/AEI/10.13039/ 501100011033, by the Agencia de Gestió d'Ajuts Univesitaris i de Recerca (AGAUR) through 2021-SGR-01470, and the CERCA Programme/Generalitat de Catalunya. B. R. was supported by "Grant TED2021-132513B-I00 funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR".
- **[1] Aesthetics Without Semantics**, C.A. Parraga, M. Muñoz Gonzalez, O. Penacchio, B. Raducanu, and X. Otazu, Submitted Computer Vision and Image Understanding (2024). **[2] StyleGAN2-ADA:** T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila. Training generative adversarial networks with limited data. Advances in neural information processing systems, 33:12104–12114, 2020. **[3] Cosine Similarity**: M. Islam, H. Zunair, and N. Mohammed. CosSIF: Cosine similarity-based image filtering to overcome low inter-class variation in synthetic medical image datasets. arXiv preprint arXiv:2307.13842, 2023. **[4] t-SNE:** L. van der Maaten and G. Hinton. Visualizing data using t-SNE. Journal of Machine Learning Research, 2008.9