Computational Photography: CS 413
Project Proposals 2023

February 2023

IVRL

Below is a list of project proposals for the CS 413 course, Spring semester 2023. Clarifications about each proposal can be obtained from the corresponding supervisor TA. The deliverables explain what is expected from you to submit by the end of the semester, aside from presentations/reports.

Table of Contents

1 Aesthetics of sets of images

Fig. 1. Wall-collages are examples of sets of images that can be considered as aesthetic sets of images. Images source: amazon.com and amazon.com

Description: The existing literature on computational aesthetics (datasets and models, [2], [5], [3], [6], [1], [4]) mainly focuses on assessing the aesthetics of images individually. The goal of this project is to extend those works to sets of images: What makes different images fit well together or not? Is a set of pleasant images always a pleasant set of images? Are images in a pleasant set of images also pleasant individually?
Tasks: As a starter point, you will assess the aesthetics of a set of images by making a collage out of those images and feed it to a pretrained aesthetics model (designed for single images), e.g., [2], [5]. You will evaluate this technique by gathering some sets of images that should be predicted as pleasant/well-fitting-together (e.g., wall collages, illustrations inside the same board game or inside the same illustrated book, letters that belong to the same font/typeface, screenshots from the same movie, etc.) and non-pleasant/non-fitting (e.g., random images, letters that belong to the different fonts, a single duplicated image, etc.). Then you will have to propose new methods and train your own models to assess the aesthetics of sets of images and find clues about what makes a set of images pleasant or not.
Prerequesites: Basics of Machine Learning. Basics of Python programming.
Learning objectives: Computational aesthetics, Machine Learning, Image classification.
Deliverables: Code, well cleaned up and easily reproducible. Written Report, explaining the literature and steps taken for the project. Data and datasets that have been used.
Supervised by: Martin Nicolas Everaert (martin.everaert@epfl.ch)

References

1. Lore Goetschalckx, Alex Andonian, Aude Oliva, and Phillip Isola. Ganalyze: Toward visual definitions of cognitive image properties. In *Proceedings of the ieee/cvf international conference on computer vision*, pages 5744–5753, 2019.
2. Simon Hentschel, Konstantin Kobs, and Andreas Hotho. Clip knows image aesthetics. *Frontiers in Artificial Intelligence*, 5, 2022.
3. Phillip Isola, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. What makes an image memorable? In *CVPR 2011*, pages 145–152. IEEE, 2011.
4. Kuan-Chuan Peng, Tsuhan Chen, Amir Sadovnik, and Andrew C Gallagher. A mixed bag of emotions: Model, predict, and transfer emotion distributions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 860–868, 2015.
5. Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE transactions on image processing*, 27(8):3998–4011, 2018.
6. Katja Thömmes and Ronald Hübner. Instagram likes for architectural photos can be predicted by quantitative balance measures and curvature. *Frontiers in Psychology*, 9:1050, 2018.

2 TimeWarp: How would this scene look in 100 years?

Description: The goal of this project is to generate a synthetic dataset using state-of-the-art diffusion models and to train an image-to-image translation model. In particular, you will focus on transforming a photograph of a scene into a new photograph of what this scene might look like after 100 years.
Tasks:
1. Synthetic dataset generation (simplified pipeline, originally from [1]): You will use text-guided diffusion models (Stable Diffusion [3]) to generate various photographs. You will also use a Cross Attention Control method (Prompt-to-prompt [2]) to generate the same photographs as they might look like if the scenes were abandoned for 100 years.
2. Image-to-image Translation training: You will then use your synthetic dataset to train a deep model (e.g. a U-Net [4]) that takes as input an image and outputs what it might look like in 100 years.
Possible variants: Several teams can work on this project. We can modify the task ("if it was abandoned for 100 years") into many variants ("if it was sunny that day", "if the scene was happening in the Middle Age", "if we were on the Moon", "in the theme of Christmas").
Prerequesites: Basics of Machine Learning. Basics of Python programming.
Learning objectives: Denoising Diffusion Models, Cross Attention, Auto-Encoders, Image-to-image Translation.
Deliverables: Code, well cleaned up and easily reproducible. Written Report, explaining the literature and steps taken for the project. Data and datasets that have been used.
Supervised by: Martin Nicolas Everaert (martin.everaert@epfl.ch)

References

1. Tim Brooks, Aleksander Holynski, and Alexei A Efros. Instructpix2pix: Learning to follow image editing instructions. *arXiv preprint arXiv:2211.09800*, 2022.
2. Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Prompt-to-prompt image editing with cross attention control. *arXiv preprint arXiv:2208.01626*, 2022.
3. Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.
4. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

3 Text to Photomosaic

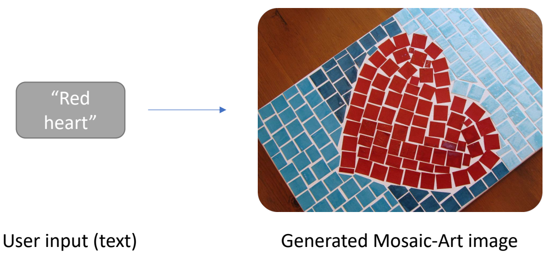

Fig. 2. Mosaic-Art synthesis from text, example of expected input/output for step 2. Image source: blogspot.com

Description: The goal of this project is to create a simple Text to Photomosaic model. Your method will not require any training, as the generation of the Photomosaic pictures will be optimization-based.
Tasks:
1. Differential rendering of Mosaic-Art: To implement the next step, you will first need to implement a differentiable renderer for Mosaic-Art. From a set of tiles (each with attributes: color, position, orientation, etc.), you must be able to render the Mosaic-Art image and compute the gradient of the pixel values with respects to the tiles attributes (color, position, orientation, etc.). You will mainly use diffvg [2] for this step.
2. Mosaic-Art synthesis from text: You will implement an optimization-based method to generate the Mosaic-Art from text. The main component of your optimization will optimize the attributes of the tiles in the Mosaic-Art to maximize the CLIP similarity [3] between the Mosaic-Art picture and the text. For this step, you will find most inspiration from CLIPdraw [1], but you have to optimize tiles instead of strokes.
3. Image retrieval for Mosaic-Art to Photomosaic: Once you have your Mosaic-Art images, you will find one photograph for each tile to obtain Photomosaics.
Prerequesites: Basics of Machine Learning. Basics of Python programming.
Learning objectives: Image Synthesis Through Optimization, Text and Image multi-modality, Differentiable renderers.
Deliverables: Code, well cleaned up and easily reproducible. Written Report, explaining the literature and steps taken for the project.
Supervised by: Martin Nicolas Everaert (martin.everaert@epfl.ch)

References

1. Kevin Frans, Lisa B Soros, and Olaf Witkowski. Clipdraw: Exploring text-to-drawing synthesis through language-image encoders. *arXiv preprint arXiv:2106.14843*, 2021.

2. Tzu-Mao Li, Michal Lukáč, Gharbi Michaël, and Jonathan Ragan-Kelley. Differentiable vector graphics rasterization for editing and learning. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 39(6):193:1–193:15, 2020.
3. Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

4 From RGB to NIR with Style Transfer

Fig. 3. Left: RGB channels of RGB-NIR images. Right: NIR spectral bands of RGB-NIR images. Image source: RGB NIR Scene dataset.

Description: While RGB imaging captures data only for red, green and blue colors, multispectral imaging captures data at various different wavelengths. In this project we focus on RGB-NIR images. The goals of this project are to create models that predict a possible NIR image from only the RGB channels of the image, and analyze the results.
Tasks: You will train Image-to-Image Translation models to predict the NIR image from the RGB image, using a RGB-NIR dataset [1]. You will analyze your model in terms of multispectral imaging. Based on your analysis (e.g., maybe the model fails on vegetation), you will improve your model with additional pretrained models (e.g., segmentation models).
Prerequesites: Basics of Machine Learning. Basics of Python programming.

Learning objectives: Multispectral imaging, Auto-Encoders, Image-to-image Translation, Style Transfer.
Deliverables: Code, well cleaned up and easily reproducible. Written Report, explaining the literature and steps taken for the project.
Supervised by: Martin Nicolas Everaert (martin.everaert@epfl.ch)

References

1. M. Brown and S. Süsstrunk. Multispectral SIFT for scene category recognition. In *Computer Vision and Pattern Recognition (CVPR11)*, pages 177–184, Colorado Springs, June 2011.