

AUGMENTify: One stop solution for generative image and video augmentation.

Naresh Kumar Devulapally, DeeKay Goswami

June 9, 2023

Title:

AUGMENTify: One stop solution for generative image and video augmentation

Short Summary:

This document is the proposal for the CSE676 Deep Learning course project during the Summer 2023 semester instructed by Professor Alina Vereshchaka. Towards this project, we propose AUGMENTify aimed at providing all the augmentation requirements in one place. We have divided the project in 3 phases, each getting progressively difficult.

Methodology:

Phase 1

This phase includes the implementation of image augmentation for any uploaded dataset. We provide the following image augmentations and ability to download the augmented dataset. Augmentations include:

- Image Padding
- Image Resize
- Center Crop
- Grayscale
- Random Color Jitter
- Gaussian Blur
- Random Perspective
- Random Rotation
- Random Affine Transform
- Elastic Transform
- Random Crop
- Random Resized Crop
- Random Invert
- Random Posterize
- Random Solarize
- Random Adjust Sharpness
- Random Equalize
- Auto Augment
- Random Augment
- AugMix
- Random Horizontal Flip
- Random Vertical Flip

We plan to implement each of the above listed augmentations utilizing PyTorch documentation and refactor the code to work with **Flask** backend. Planned deadline for this task is **10th June 2023**.

Phase 2

Once the image augmentation part of the project is completed, we proceed to **Text query based image retrieval**. This would enable the user an option to retrieve images from the uploaded or existing datasets based on a text query input. To complete this, we plan to use a visual-language model to generate descriptions for each image in the dataset then use **cosine similarity** to find the images whose descriptions are close to the input text query.

We plan to make the search similar to image search using the Google search engine. However, this search would be throughout the selected dataset. Our plans include to store the text-based image retrieval model to yield much faster results to future users. We are still looking for potential cloud storage options to store the trained models. Planned deadline for Phase 2 is **20th June 2023**.

Search results should look something like this:

Query: "Few images of red cars."

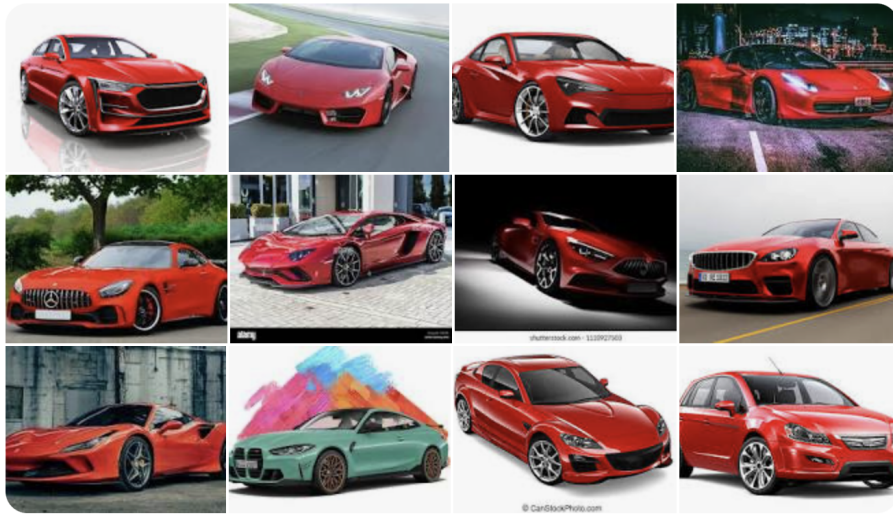


Figure 1: Text query based image retrieval (example).

Phase 3

Final phase of the project is to perform **Generative Image editing using text prompt**. We plan to leverage the power of pretrained Generative Adversarial Networks including, StyleGAN [KLA18], EditGAN [LKL⁺21], StyleGAN2 [KLA⁺19] and other Generative models based image editing works.

Firstly, we look forward to changing simple attributes such as color of the objects retrieved in Phase 2 of the project. The text query can be a simple sentence such as **"Change the color of the cars from red to blue"**.

We believe that Phase 3 is the most difficult part of the project given the time constraints. Hence, we try to complete this part by the end of the course during Summer 2023.

Upon completion of the Generative Image editing in the Phase 3. We plan to use the repo VidAug, to produce video augmentation to users.

Finally as an extension to our work, we plan to deploy the project as a web application using Flask backend and ReactJS frontend. We have a plan and created wireframes of the final application as shown in the supplementary material in the last page.

Evaluation:

The evaluation of the project is the working of each feature in AUGMENTify. We plan to utilize flask to create endpoints for every feature in the application and test the working of each feature.

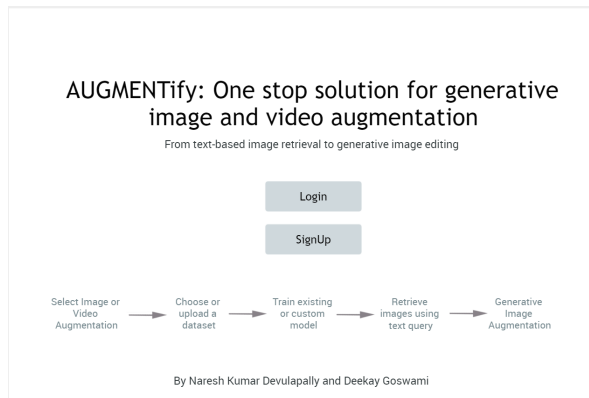
Dataset:

AUGMENTify is aimed at working on any dataset. However, training Generative editing might make the use of pretrained models on various datasets. We are still looking into the details of various datasets such as celebA, SVHN and so on.

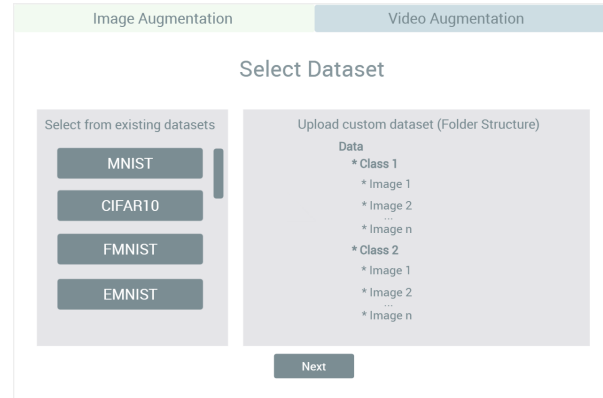
References

- [KLA18] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. *CoRR*, abs/1812.04948, 2018.
- [KLA⁺19] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. *CoRR*, abs/1912.04958, 2019.
- [LKL⁺21] Huan Ling, Karsten Kreis, Daiqing Li, Seung Wook Kim, Antonio Torralba, and Sanja Fidler. Editgan: High-precision semantic image editing. *CoRR*, abs/2111.03186, 2021.

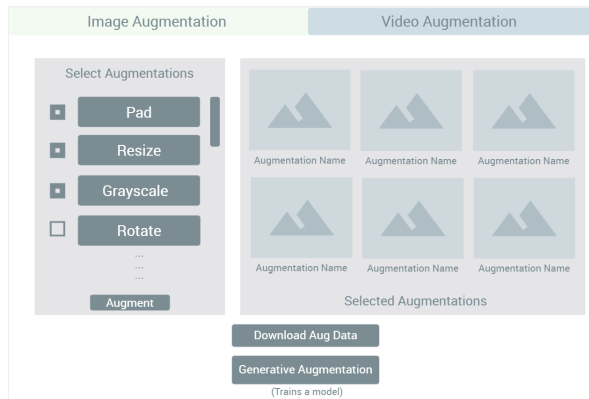
Supplementary Material



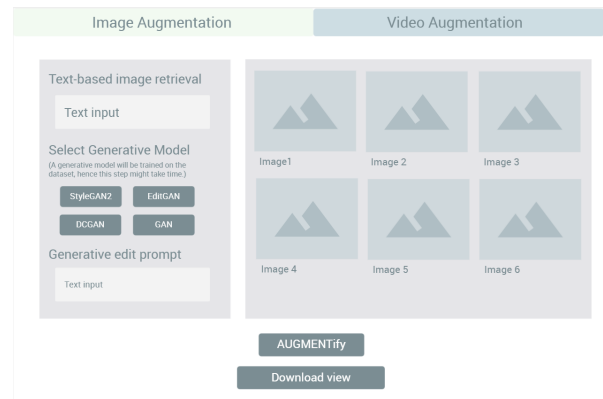
(a) Login Screen



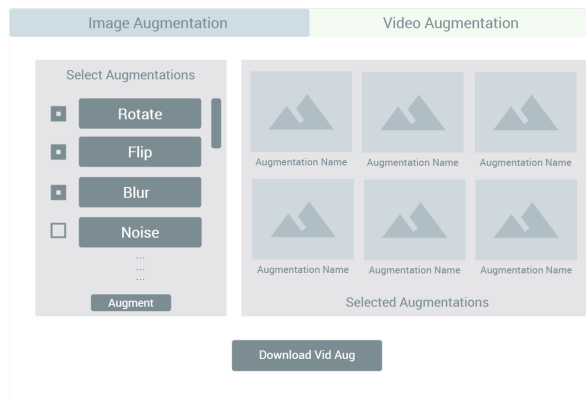
(b) Select dataset or upload



(c) Image Augmentation



(d) Generative Editing



(e) Video Augmentation

Figure 2: Wireframes for AUGMENTify Web App