Final Project

DL4CV @ WIS 2021

1 Introduction

The final assignment is implementing a Deep Learning project of your own. You are expected to leverage the knowledge and techniques taught in this course to perform a task in the Computer Vision domain. You may either choose a project from the default project list (see Section 6) or pursue an idea of your own. We encourage you to come up with an original idea for your project. You are particularly welcome to choose tasks that will benefit your lab's research or follow your personal interests. You will conclude your work by submitting a Project Report. The details of the Project Report will be announced further during the semester. Also, we plan to hold a Project Presentation Day where you will present your work.

1.1 Important Notes:

- Work in groups of 2-3 students.
- Group Members & Project Proposal Date: June 17^{th}
- Project Milestone Date: July 1^{st}
- Project Submission Date: August 5^{th}
- Project Presentation Date: TBD

2 Project Proposal (Due: June 17^{th})

All groups are required to submit a Project Proposal Paper. The project proposals enable us to monitor the feasibility of your projects, and point you toward relevant directions and resources.

2.1 Deliverable

The Project Proposal should be submitted in PDF format, the file name should be `project_proposal_member1_member2_member3.pdf` where each member is a name of a member of the group. The file may contain **up to 2 pages**.

In your proposal make sure to contain the following sections:

1. **Group Members:** List all the members of the team and their affiliation (faculty, degree and PI).

2. **Abstract:** Describe the task you aim to solve. Provide a short explanation of the techniques you plan to use (e.g. CNN, GAN, LSTM, self-supervision, etc.) and how to combine them into a single framework (e.g. describe your training & inference process).

3. **Data:** Characterize the dataset you wish to use. Namely, state the number of samples, size in memory, availability of the dataset and its contents. Include some data samples if possible.

4. **Resources:** Mention any special hardware resources (CPU, GPU) you will use to develop your project. We recommend using Google Colab as in previous assignments in the course, which provide a single K80 GPU.

5. **Challenges:** Briefly mention any technical or conceptual challenges you foresee and how you plan to overcome them.

3 Project Milestone (Due: July 1^{st})

The purpose of the project milestone is to enable us to ensure that you are on track with implementing your proposed project. We want to see that you have all the components and "moving parts" in place and it is clear to you how to proceed.

3.1 Deliverable

The Milestone report should be submitted in PDF format, the file name should be `milestone_member1_member2_member3.pdf` where each member is a name of a member of the group. The file may contain **up to 2 pages**.

The milestone report is very similar to the proposal:

1. **Group Members**.

2. **Abstract**.

3. **Data:** Describe the datasets that you are using. Have you been able to get the data you expected? If not, what are you missing and how is it going to affect the final outcome? Include some data samples if possible.

4. **Method:** Briefly describe your approach.

5. **Preliminary Results:** These are only *preliminary* indications that all "moving parts" are in place.

6. **Rising Flags:** Have you encountered, or anticipating, any issues that might prevent you from completing the project?

4 Project Submission (Due: August 5^{th})

4.1 Deliverable

A 5 minutes video in which you describe your project. The description must include the following parts:

1. **Introduction:** Describe the problem you are working on, why it's important.
2. **Related work:** Put your work in context, what has been done, how your project relates to previous work. Was your approach inspired by other works?
3. **Method:** How you tackled the problem, what data have you used and what tools and approaches taught in the course you used.
4. **Results:** Describe your results.
5. **Conclusions:** What insights and lessons have you learned from working on this project?

You should submit two files: a pptx file with the slides and an mp4 video recording of that presentation.

5 Project Presentation (Due: TBD)

You would be required to briefly present your work to all other students in the course. We expect this event to be a mini-conference showcasing your efforts to your peers.

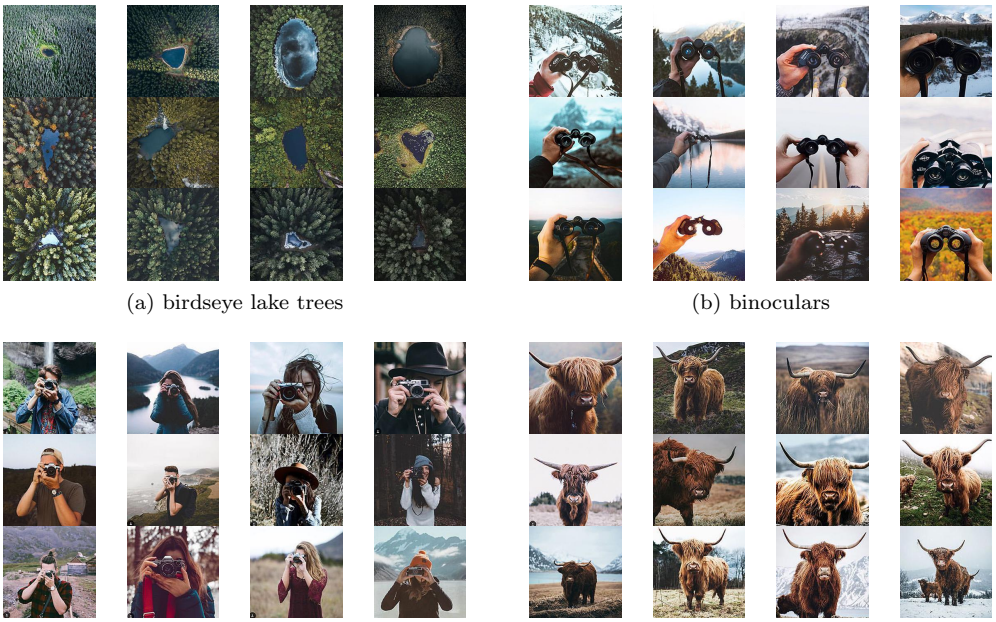
6 Default Projects List

6.1 Default Project 1 - Instarepeat

In this project you will solve the grouping task of the `Instarepeat` dataset.

6.1.1 Dataset

The `Instarepeat` dataset contains images from the popular Instagram account with the same name. The account features images posted in popular travelers' Instagram accounts that have a common appearance. Each post contains a grid composed of 12 thumbnails of images with similar content and visual appearance. We grouped all grids into groups marked with some label (see Figure 1). Some images from different grids belong to the same group, so a group may contain more than 12 images. In total there are around 300 grids, belonging to roughly 120 labels. Each grid contains 12 images, thus there are approximately 4K images.



(a) birdseye lake trees (b) binoculars

(c) person camera (d) wildcow

Figure 1: Examples for Instarepeat grids.

We split the dataset into a training and testing sets. Deciding how to do validation on the training set is up to you. The testing set contains some grids from the same groups as the training set and some groups that are unique for the testing set. You MAY NOT use the testing set for any purposes besides evaluation.

6.1.2 Task

Your task is to reconstruct the groups. In other words, assign each image to some group. This should be done in an unsupervised manner. Namely, your algorithm may not be aware to the original labels of the test-set, they are given for evaluation purposes alone.

The task will be evaluated using two metrics - Adjusted Random Index (abbrev. ARI) and Normalized Mutual Information (abbrev. NMI). For formal definitions see NMI definition and ARI definition. You should report the value of these metrics on the testing set. The ARI and NMI values on the testing set should be significantly above 0.44 and 0.75 respectively, which is achieved using a baseline of a pretrained backbone with no extra training. You may assume there are 45 groups in the test set.

In addition, you are most encouraged to show insightful visualizations of your results. For example, show the t-SNE visualization of your grouping with

comparison to the original groupings. Also, you may show interesting failure cases and try to interpret them.

6.1.3 Resources

We recommend working in Google Colab. The GPU resources provided by them should suffice.

6.1.4 Approaches

You may approach this task in any way you see fit, as long as you train some network in your solution. You may decide to approach this task in a supervised manner - using the training set's labels, or in an unsupervised manner. For your convenience we include several recommended approaches.

- Train a neural network to perform clustering on the training set. You may use the labels as ground truth clusters.
- Train a neural network in a self-supervised manner, perhaps using a contrastive loss. Then use the testing set's embeddings as inputs to some clustering algorithm (e.g. K-means).
- Train an auto-encoder to learn good representations and continue as above.
- Do Transfer Learning - applying any of the above using pretrained backbones (e.g. on ImageNet).

6.2 Default Project 2 - Spot The Fake

In this project you will distinguish real and fake images, as well as classify the fake image's origin.

6.2.1 Dataset

The dataset contains unaligned face images that belong to 5 different sources:

1. `Taming` - Images from the paper *Taming Transformers for High-Resolution Image Synthesis* by Esser et al.
2. `StyleGANv2` - Images generated from the paper *Analyzing and Improving the Image Quality of StyleGAN* by Karras et al.
3. `real` - Images from the Flicker Faces HQ (abbrev. FFHQ) dataset.
4. `ProGAN` - Images from the paper *Progressive Growing of GANs for Improved Quality, Stability, and Variation* by Karras et al.
5. `NVAE` - Images from the paper *NVAE: A Deep Hierarchical Variational Autoencoder* by Vahdat et al.

Note that `NVAE` is the only generative model which isn't a GAN. The training set contains roughly 800 images per source while the validation set contains roughly 100 images per source.

The testing set contains around 300 images from all these sources, and another extra testing set contains 300 more images from an unseen source.

6.2.2 Task

This project is divided into two tasks:

- **Real / Fake Classification** - Train a network to classify images as `real` or `fake`. Testing will be done on both testing sets.
- **Source Classification** - Train a network to classify the source of the image, out of the five aforementioned classes. Testing will be done on the first testing set alone.

In addition, we encourage you to explore visualizations that will shed light on the functionality of these networks. For example, use GradCAM to visualize dominant areas for the neural network's decision.

6.2.3 Resources

We recommend working in Google Colab. The GPU resources provided by them should suffice.

6.2.4 Approaches

You may approach this task in any way you see fit, as long as you train some network in your solutions. For your convenience we include several recommended approaches.

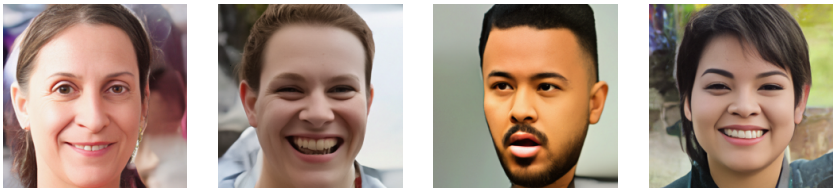
- Train a vanilla CNN network with cross entropy loss for each of the tasks.
- Train a network on patches, perhaps in a self-supervised manner.
- Train a network on high frequency features (e.g. Fourier Features) rather than the original image.
- Do Transfer Learning - applying any of the above using pretrained backbones (e.g. on ImageNet).

(a) Taming

(b) StyleGANv2

(c) real

(d) ProGAN



(e) nvae

Figure 2: Examples for images from different sources.