[Fall 2022] CS272 Final Project

October 18, 2022

Acknowledgements

1. Proposal and collaboration registration deadline: **2022/10/26 23:59:00** (week 8). No late days.
2. If you choose to propose your personal project,
   - please submit your proposal and slides in Gradescope with PDF format.
   - you are required to give a brief presentation for proposal in about week 9.
3. Plagiarism or cheat is strictly prohibited.

Final Project is an opportunity for you to apply what you have learned in class to a problem of your interest in computer vision. Here we offer you two options to build your final project.

1 Propose your personal project

You could select a topic in computer vision that you are working on and propose it as your personal project. Personal project should **only include 1 student**. But if you have a parallel collaborator, please make a statement to TA in e-mail and CC your supervisor. Proposal should be a **1-page PDF** and **slides (about 3min)** that answer the following questions:

- **Project Title**: What is the name of your project?
- **Group Members**: (if you have a parallel collaborator) What are the names and ids of the students involved?
- **Problem Statement**: What is the problem you are trying to solve? Why is it interesting? What reading will you examine to provide context and background?
- **Approach**: What method or algorithm are you proposing? If there are existing implementations, will you use them and how? How do you plan to improve or modify such implementations? You don't have to have an exact answer at this point, but you should have a general sense of how you will approach the problem you are working on.
- **Data**: What dataset do you plan to use?
- **Evaluation**: How do you plan to evaluate whether your project is successful? What metric will you use? Is there some simple baseline that you plan to compare your model against?

Please submit your proposal and slides as a PDF on Gradescope. If you have a parallel collaborator, only one person on your team should submit. Please have this person add the rest of the team members to the group submission on Gradescope. Later, you will give a brief presentation for your proposal in class.

2 Choose from the suggested topics

You could choose from the list of suggested topics. Each team can be **up to 5 students**. If you choose suggested topics, you don't have to submit a proposal or slides. Suggested topics include:

1. Digital Photo Album APP
2. Pose-guided Action Quality Assessment

2.1 Digital Photo Album APP

Recently, the many album apps have launched smart album module, such as Photos app (Apple album), Baidu Netdisk, etc. For instance, the Photos app can recognize the faces of people in your photos and group them together (Fig.1). You can name the people in your photos. Also, the Photos app can recognize significant people, places and events in your library, then presents them in curated collections called Memories.

You could also design and build your own Digital Photo Album APP using what you have learned in class. Some basic functions should be included in your app, such as face recognition, image grouping, image retrieval, etc. Besides, you are encouraged to design some novel functions to enrich your app, such as vision-language model (Fig.2 (a)), facial expression recognition (Fig.2 (b)), event recognition (Fig.3), etc.

Training dataset and album photos are not specified, and you can build your dataset based on some public datasets according to your function design. To make your app more vivid, we highly recommend you to use your real photos (instead of public dataset or search engine) to build your album.

You don't have to build a real app but you are required to realize the function you have designed. But a demo platform presentation is highly appreciated. You need to specify your function design, model design, method design and other details (dataset construction, implementation details etc.) in your final report.

Hint: if you would like to use vision-language model, we recommend you try CLIP[1] to boost your performance.

The Photos app guidance:

- https://support.apple.com/en-gb/HT207023
- https://support.apple.com/en-gb/HT207103

Some related public datasets:

- https://paperswithcode.com/dataset/imc-phototourism
- https://paperswithcode.com/dataset/photosynth
- https://paperswithcode.com/dataset/memexqa
- https://paperswithcode.com/paper/mobiface-a-novel-dataset-for-mobile-face
- https://paperswithcode.com/dataset/widerperson
- https://paperswithcode.com/dataset/wider
- https://paperswithcode.com/dataset/wider-attribute-dataset
- http://shuoyang1213.me/WIDERFACE/

References:

[1] [ICML 2021] CLIP: Learning Transferable Visual Models From Natural Language Supervision

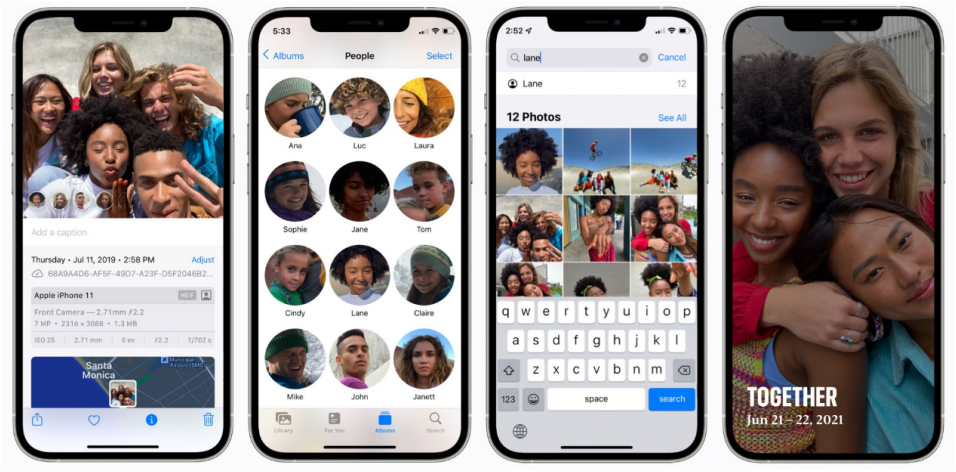
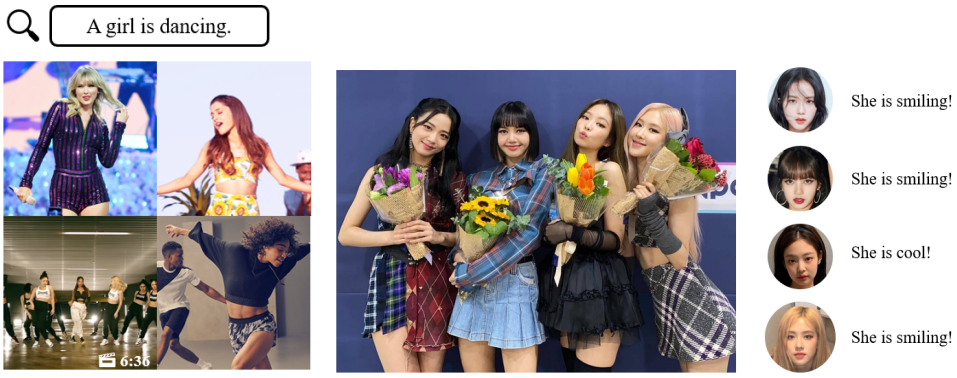


Figure 1: The Photos app.



(a) Vision and language (b) Facial recognition.

Figure 2: Some novel ideas.

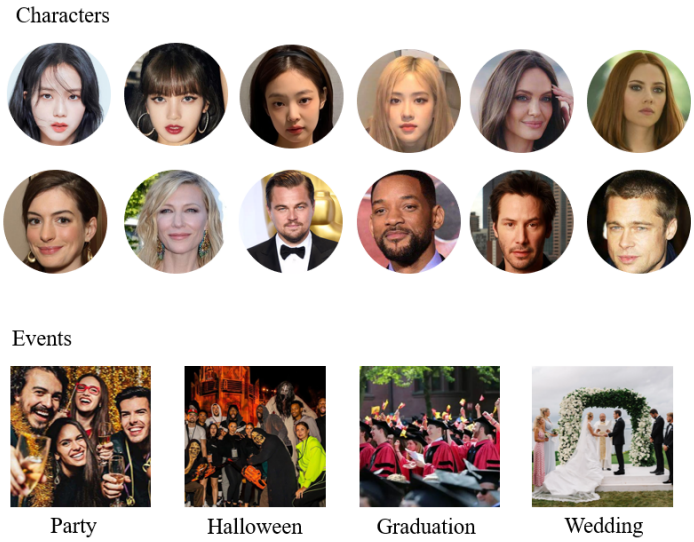


Figure 3: Event recognition.

2.2 Pose-guided Action Quality Assessment

Competitive sports video understanding has become a hot research topic in the computer vision community. As one of the key techniques of understanding sports action, Action Quality Assessment (AQA) has attracted growing attention in recent years. In the 2020 Tokyo Olympic Games, the AI scoring system in gymnastics acted as a judge for assessing the athlete's score performance and providing feedback for improving the athlete's competitive skill, which reduces the controversies in many subjective scoring events, e.g., diving and gymnastics.

You can do this like below, of course, other methods are allowed. Just remember all check points should be done.

Part 1 Reading Preparing Dataset. In this project, we use FineDiving[1] as our dataset for training and testing. Download the FineDiving dataset and understand the format, try to read the image data and labels. Checkpoints: Please write a function to randomly display an image and its corresponding label from the dataset.

Part 2 Pose Estimation. Pose estimation from a single monocular RGB image aims to simultaneously isolate and locate body joints of multiple person instances. It is a fundamental yet challenging task with broad applications in action recognition, person Re-ID, pedestrian tracking, etc. Here you are requtested to select one of SOTA pose estimation network to estimate the skelecton representation of action.

Checkpoints: You need to write down the detail of your chosen method and show the resulting visualization result of one of action video frames.

Part 3 Action Quality Assessment. In this part, you are requested to designed an action quality assessment method to predict the action quality score of the query video based on the predicted pose. The action quality assessment can be formulated as a regression problem that predicts the action quality score of the query video.

Checkpoints: (1) Describe implement details of your method and experiment settings. (2) Shows the experimental results of your approach, trained and evaluated on the FineDiving dataset. Results should include at least two following metrics: Spearman's rank correlation and Relative l2-distance.

Part 4 Bonus (Optional). Skelecton representation only ulitize lower dimension feature representation, which inspire us to ulitize more molidaty. In this part, your are encouraged to develop more robust assessment method.

Checkpoints: (1) Describe implement details of your method and experiment settings. (2) Shows the experimental results of your approach, trained and evaluated on the FineDiving dataset. Results should include at least two following metrics: Spearman's rank correlation and Relative l2-distance.

Reference

[1] [arXiv 2204] FineDiving: A Fine-grained Dataset for Procedure aware Action Quality Assessment