Research Plan

Hatting Face

January 19, 2024

Abstract

We are introducing a new to-build amazing tool, adding Christmas' gadgets to group photo. Based on our in-depth analysis of the most buzzed-about photos on the Internet, we have noticed a pattern in the addition of gadgets based on people position. This document is our first try of research plan for the team. This project is going to last a year. You are pleased to suggest ideas and recommendations of any kind!

1 Introduction

We first describe the task in order to have a macro view of our next few months. The tool must follows the next conditions:

- The person furthest from the camera should have a santa hat placed on their head
- The closest person should have a red nose and reindeer antlers placed on their head.If this is the same person getting the santa hat, then the hat or antlers are randomly chosen.
- At least one person should be holding a candy cane if an arm is in shot
- It has to be highly realistic

2 How we are going to tackle this challenge ?

2.1 Object Detection

Our first approach and our current demonstration is relying on object detection. We could detect the faces, the nose and arms, if there is any, of person in the image and then adding gadgets. In order to know who was the closest and the furthest we would be computing the area of the head and trying to order them, the smallest is the furthest, the biggest the closest.

We could focus our research on having the best detection tool but it would be not really interesting for the team and the detection part could be a really hard thing to handle in a lot of cases, for example a picture where people does not have a straight position. Indeed, object detection still have struggled with finding the good orientation. Moreover, it will not be highly realistic since we are unable to close a hand on an object with object detection, and this what we will need to be realistic with the candy. Therefore, we drop this idea for now but would be happy to heard counter-arguments!

2.2 Segmentation

The segmentation idea was also an idea, however we are encountering the same problem as with object detection.

2.3 Diffusion model and Flow Matching

Recent research work has introduce several diffusion model in order to modify an image with a prompt. These kind of approaches are multi-modal ones and usually enable the model to better understand the specific task. We are going to explore these two approachs for our product!

3 Plan

3.1 Study of Diffusion Models & Flow Matching Techniques

- In-depth literature review and testing of diffusion models and flow matching for images. Example of papers [1][2].

3.2 Definition of metrics and datasets

- For high realistic measurement a first idea would be to evaluate it with a GAN discriminator.
- An other idea would be to substrate the input image from the generated image. A perfect generation would be to have only the gadgets on the remaining image.
- Looking at [1] paper, reproduce their method to create a dataset. Creation of a dataset for our specific task.

3.3 Common failures

- Common failures assessment based on testing feedback.
- We expect failures for photo taken from a distance
- Handling pictures with multiple people and ensuring that no gadgets is overlapping or incorrect placement can be challenging.
- People with hats.
- People in the pictures have unusual head positions, like looking away or tilting their heads at extreme angle

3.4 Other Features

- Theme variety (Easter for example)
- Make the gadget adapt to the person emotion face
- Colors gadgets variation

3.5 System Optimization

- Efficient integration of model into a unified system.
- Optimization for real-time processing and accuracy.

3.6 User Experience Testing

- Conducting user testing sessions to evaluate usability and output quality.
- Feedback incorporation for system refinement.

References

[1] InstructPix2Pix: Learning to Follow Image Editing Instructions, Brooks, Tim and Holynski, Aleksander and Efros, Alexei A, 2022

[2] Flow Match For Generative Modeling, Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, Matt Le, 2022.