Harvard AC215 (Shiqi Wang, Yanfeiyun Wu, Jiahui Zhang, Wanying Zhang)

Statement of Work

Problem Statement

The primary objective of our project is to develop a Deep Learning model capable of interpreting an image of a crochet product and generating a corresponding set of pattern instructions.

Crochet has long been a popular and therapeutic activity for many, offering not only a creative outlet but also a relaxing and rewarding pastime. Our group of four has developed a deep passion for crochet over time, but despite our enthusiasm, we have consistently encountered challenges in accessing reliable and comprehensive instructional patterns for products we find interesting. Whether it's a unique design we come across online or an object created by others that catches our eye, finding suitable texture pattern instructions that match the design can be a frustrating process due to the scarcity of resources.

In the course of our research, we discovered a tool that could transform written instructions into visual 3D models, allowing users to preview how the crochet item would appear once completed (website: https://new.weishougong.cn/crochet). However, we noticed a significant gap in the field—there has been no accessible tool or research focusing on generating crochet pattern instructions based on images of finished crochet products. This is where we see an opportunity to innovate and address a genuine need in the crafting community.

Minimum Components for a Good Project

- **Large Data**: Collection and utilization of a varied 2D crochet image and corresponding textual pattern instructions.
- **Scalability**: Ability to handle multiple users querying textual patterns using their uploaded images.
- **Complex Models**: Use of deep learning-based computer vision models for feature extractions, and use of multi-modal vision and language models for textual pattern generations.
- **Computationally Expensive Inference:** Implement effective models that can shorten crochet textual instruction searching time.

Objectives

- Collect and preprocess a diverse dataset of crochet products images.
  - **Data Sources**: The dataset will be sourced from a publicly accessible website that offers free textual instructions for a wide range of crochet goods. (https://www.yarnspirations.com/)
  - **Data Attributes**: The dataset will comprise images of various crochet items alongside their corresponding detailed textual instructions, providing a comprehensive overview of each product.
  - **Data relevance**: This dataset encompasses a broad spectrum of crochet goods, including popular categories such as toys, cardigans, and scarves, ensuring a diverse and representative collection for analysis or modeling.
  - **Data Quality:** The images in the dataset may only capture a partial view of the crochet goods, as they represent 2D depictions of inherently 3D items.
- Develop a computer vision model and multi-modal vision and a language model to transform crochet good images into textual patterns.

- Design an intuitive and user-friendly frontend.

Learning Emphasis

The project will focus on employing an image feature extraction model to extract features from the input crochet product image and a multi-modal vision and language model to map the image features to generate a crochet pattern instruction.

Application Mock Design

The application will feature three main interfaces:

- Interfaces for photo upload and camera for capturing or uploading images of knitted goods.
- A pattern output interface to display the pattern instructions based on the knit goods.
- An interface for pattern saving.

Research and Development

We will review and reference open-source blog posts on crochet projects images for feature extractions and multi-modal vision and language models for the pattern instructions generation.

Fun Factor

This project tackles an important need in the crafting world, blending our shared passion for crochet and machine learning. By combining these two areas, we're creating a tool that inspires creativity and makes the art of crochet more accessible to everyone.

Limitations and Risks

- Obtaining enough accurate crochet product images and corresponding texture instructions manually from public websites is time-consuming.
- Possibility of unbalanced data - ensuring the dataset is diverse enough to cover different types of crochet products is challenging.
- Risk of overfitting - the model may perform well on training data and generate excellent instructions for seen crochet products, but produce poor instructions for unseen ones.
- Even if the model can generate instructions, verifying the instructions are correct and usable is hard and time-consuming.
- The complexity of the model may result in high computational costs.

Milestones

Data collection and preprocessing: [October 1, 2024]

Computer vision model development: [October 18, 2024]

Backend implementation: [October 28, 2024]

Frontend development: [November 10, 2024]

Final testing and deployment: [November 20, 2024]