

Design Document

Hand-Drawn Animals Recognition Using Voice Interface

Sven Ludwig, Hoan Vu

University of Applied Sciences Bonn-Rhein-Sieg

Visual Computing and Games Technology

14. November 2023

1 Introduction

Children have a natural love for artistic expression, and among their favorite subjects are animals. What if we could take this creativity a step further and allow a child to create stories around their own hand-drawing favorite animals by interacting with a computer to shape their artwork? This project helps to spark their creativity with the capabilities of computer vision through voice interface technology. The primary aim of this project is to develop a tool that recognizes hand-drawn animal images, engages in a friendly dialogue with children, and encourages interactive learning in a fun and educational way.

2 Objectives

The main objectives of this project are as follows:

- To create a tool that takes as input a photo of a hand-drawn animal image.
- To implement an animal recognition system that classifies the drawings into specific animal classes, e.g. fox, owl, deer, cat, crocodile, and spider.
- To enable a child-friendly interaction between the tool and the child, adhering to the principles of Advanced UI design.
- To log and record instances where children present drawings of animals not covered by the initial set of recognized classes.
- To provide a warm welcome message when the tool is initiated and a friendly farewell message when the interaction session concludes.

3 Methodology

In order to achieve the outlined objectives, the following methodology will be employed. Note that this project won't cover the image classification part:

- Integration of the voice interface will necessitate software capable of speech recognition and generation, e.g. using technologies like Google Speech Recognition and Text-to-Speech (TTS) APIs.
- Potentially, the development of the interactive dialogue with children will involve small-scale natural language processing (NLP) techniques and user interface design principles to create a friendly and engaging experience.

3.1 Hardware

- Camera or image input devices (webcam, smartphone camera) for capturing and uploading hand-drawn animal images.
- Microphone and speaker system for the voice interface, ensuring clear and accurate communication with the child.

3.2 Software

- The first party software is implemented using python.
- The applied build system is poetry.
- The code is hosted on GitHub.

3.3 Potential voice solutions

- CMU Sphinx (PocketSphinx) by Carnegie Mellon University is an open source lightweight library written in C, with python API via pocketsphinx package. <https://cmusphinx.github.io/>
- DeepSpeech by Mozilla is a more complex powerful deep learning model, written with Tensorflow. <https://github.com/mozilla/DeepSpeech>

4 Timeline

14.11. Project setup

28.11. Dataset curation, initial preprocessing, architecture overview

05.12. Basic recognition of drawn animals

19.12. Text-to-Speech asking the first question of the dialogue

09.01. Dialogue version 1

16.01. Dialogue version 2 and technical improvements

23.01. MVP done.