CAT FACE IDENTIFICATION USING CLASSICAL COMPUTER VISION TECHNIQUES

A Project Report

This project presents a classical computer vision approach for cat face identification using handcrafted features and traditional machine learning techniques. Facial landmark annotations are used to localize cat face regions, enabling accurate extraction of positive samples without relying on learned detection models. Background samples are generated from non-overlapping image regions to formulate a binary classification problem.

Histogram of Oriented Gradients (HOG) features are extracted to represent structural characteristics of facial regions, and a linear Support Vector Machine (SVM) classifier is trained to distinguish between cat face and background regions. Experimental results demonstrate that the proposed system achieves high classification accuracy while maintaining interpretability and computational efficiency. The project highlights the effectiveness of classical vision pipelines and reinforces their relevance as a foundation for modern computer vision systems.

1. INTRODUCTION

Computer vision is a field of artificial intelligence that enables machines to interpret and understand visual information from images and videos. From unlocking smartphones using face recognition to analyzing medical images, computer vision plays a crucial role in many real-world applications. At the core of several of these applications lies a fundamental task: face detection.

While human face detection has been extensively researched, detecting animal faces presents a different set of challenges. Animals, especially cats, exhibit large variations in pose, facial orientation, fur texture, lighting conditions, and background clutter. Unlike human faces, cat faces do not always follow a rigid structure, making automatic detection more complex and interesting from a computer vision perspective.

In recent years, deep learning approaches have become the dominant solution for face detection tasks. However, classical computer vision techniques remain highly relevant, particularly for understanding the foundations of visual recognition systems. These methods rely on handcrafted features and traditional machine learning algorithms, offering advantages such as interpretability, lower computational requirements, and easier debugging compared to deep neural networks.

This project focuses on cat face identification using classical computer vision techniques. By utilizing facial landmark annotations to localize face regions, extracting meaningful image features, and applying a traditional classifier, the project demonstrates that accurate cat face identification can be achieved without relying on deep learning models. The work emphasizes a structured and explainable approach to face detection, making it suitable for educational purposes and environments with limited computational resources.

2. Literature Survey Content

Early research in face detection focused on classical computer vision techniques that relied on handcrafted features and traditional classifiers. One of the most influential works in this area was the Viola–Jones framework, which introduced Haar-like features combined with cascade classifiers for rapid face detection. This method demonstrated that simple intensity-based features could effectively capture facial structures when paired with efficient learning algorithms.

Subsequent research explored more robust feature representations to handle variations in pose, lighting, and appearance. Histogram of Oriented Gradients (HOG) emerged as a powerful feature descriptor that captures local edge orientations and shape information. By representing the distribution of gradients in localized regions, HOG features provide a compact yet expressive description of object structure. This approach has been successfully applied to various object detection tasks, including face and pedestrian detection.

Support Vector Machines (SVMs) have commonly been used alongside handcrafted features due to their strong generalization capabilities in high-dimensional feature spaces. Linear SVMs, in particular, offer a good balance between accuracy and computational efficiency, making them suitable for real-time and resource-constrained applications.

Although deep learning techniques now dominate face detection research, classical methods continue to play an important role. They offer transparency in decision-making, require less data, and serve as a strong foundation for understanding modern computer vision systems. This project builds upon these classical ideas to design an interpretable and efficient cat face identification system.

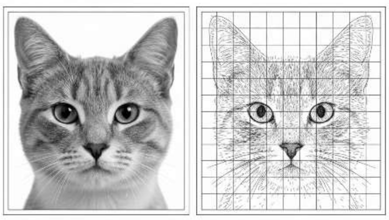


Figure 1: Illustration of handcrafted features used in classical face detection, highlighting gradient-based representations for capturing facial structure.

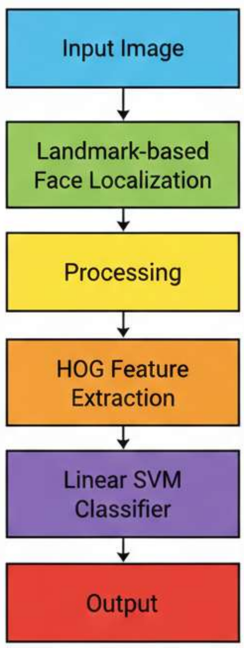
3. Problem Statement

The objective of this project is to design and implement a classical computer vision system for identifying cat face regions in images. The problem is formulated as a binary classification task, where a given image region is classified as either a cat face or a background region.

The system uses facial landmark annotations to localize cat face regions, generates background samples from non-face areas, and applies handcrafted feature extraction techniques to represent image structure. A traditional machine learning classifier is then trained to distinguish between face and non-face regions based on these features.

The focus of the project is to demonstrate an interpretable and efficient approach to cat face identification without relying on deep learning models, while achieving reliable performance on real-world images.

4. METHODOLOGY AND IMPLEMENTATION



4.1 Dataset Description

The Crawford Cat Dataset is used in this project. The dataset contains images of cats along with corresponding annotation files. Each annotation file includes facial landmark coordinates representing key points such as the eyes, nose, and mouth. These landmarks provide precise information about the location and structure of the cat face within the image.

Unlike many object detection datasets, the dataset does not directly provide bounding boxes for the face regions. Therefore, facial landmarks are used to derive the face region in a reliable and consistent manner.

4.2 Face Localization Using Facial Landmarks

To localize the cat face, the facial landmark coordinates provided in the annotation files are parsed for each image. The minimum and maximum x and y coordinates of these landmarks are computed to form a bounding box around the face. Additional padding is added to the bounding box to ensure complete coverage of facial features.

This landmark-based localization approach allows accurate extraction of the face region without using any learned detection model and ensures consistency across different images.

4.3 Background Sample Generation

To train a binary classifier, background samples are required in addition to face samples. Background regions are generated by randomly sampling patches from the image that do not overlap significantly with the face region.

The Intersection over Union (IoU) metric is used to measure overlap between the face bounding box and candidate background regions. Only patches with minimal overlap are selected as background samples. This strategy ensures that background samples do not contain facial information and helps the classifier learn a clear distinction between face and non-face regions.

4.4 Image Preprocessing

All extracted image regions are converted to grayscale to reduce computational complexity and remove color dependency. Each region is then resized to a fixed resolution of 64×64 pixels. This normalization step ensures uniform feature extraction and consistent input dimensions for the classifier.

4.5 Feature Extraction Using Handcrafted Features

Histogram of Oriented Gradients (HOG) features are extracted from each preprocessed image region. HOG captures local gradient orientation patterns, which are effective in representing structural characteristics such as edges and facial contours.

Cat faces exhibit consistent gradient patterns around regions such as the eyes, nose, and ears, making HOG a suitable choice for this task. Each image is represented as a fixed-length feature vector derived from the HOG descriptor.

4.6 Classification Using Support Vector Machine

A linear Support Vector Machine (SVM) classifier is trained using the extracted HOG features. The classifier learns a decision boundary that separates cat face regions from background regions based on their feature representations.

Linear SVM is chosen due to its effectiveness with high-dimensional handcrafted features and its computational efficiency. The dataset is split into training and testing sets to evaluate the generalization performance of the classifier.

4.7 System Architecture Overview

The overall architecture of the proposed system follows a sequential pipeline. Input images are first processed to extract face and background regions using landmark-based localization. These regions are preprocessed and converted into feature vectors using HOG. Finally, a linear SVM classifier is used to identify whether a given region corresponds to a cat face or background.

This modular architecture ensures interpretability and allows each component to be independently analyzed and improved.

5. RESULTS AND CONCLUSION

5.1 Results

The performance of the proposed cat face identification system was evaluated using a train–test split of the prepared dataset. The classifier was trained on handcrafted HOG features extracted from both cat face regions and background regions.

The system achieved an overall classification accuracy of approximately 96 percent on the test set. The high precision and recall values indicate that the model is effective in correctly identifying cat face regions while minimizing false detections of background regions.

Qualitative evaluation further demonstrates that the classifier successfully distinguishes facial regions from background clutter under varying lighting conditions and facial orientations. Sample predictions show that the model consistently captures structural features such as eyes and facial contours, which are critical for accurate face identification.