Which Animal Are You?:
A Computational Classification of Human Facial Traits

Assignment 2-1: Group 7 (G7)

Tatsuya Yamada*[1], Yizi Wang†[1], Futa Oda‡[1], and Takaya Miki§[1]

[1]Graduate School of Information Science and Technology, The University of Osaka

July 22, 2025

Abstract

Subjective impressions such as "cat-like" or "dog-like" are commonly used to describe human faces, yet their computational basis is unclear. This study investigates whether a classical machine learning pipeline can identify these perceived traits after being trained exclusively on animal faces. Our approach utilizes principal component analysis (PCA) for feature extraction and a support vector machine (SVM) for classification on a curated three-class dataset of animal images. We first demonstrate that this model achieves high accuracy on its source domain, confirming it is well-trained. However, when tasked with classifying human faces intentionally generated to appear "animal-like," its performance declines significantly, revealing a severe domain gap and a strong classification bias. A comparative experiment with a modern deep learning model shows substantially improved performance, suggesting the limitations are characteristic of the classical feature extraction method. These findings clarify the challenges of using a global, variance-based feature space for this abstract, cross-domain task and indicate that learned, hierarchical features are more suitable.

1 Introduction

Human beings often rely on subjective, analogical descriptions to characterize facial features, such as "cat-like" or "dog-like." This study investigates whether a computational model can learn to classify these subjective traits. Our primary focus is on the challenge of domain shift: we train a model exclusively on animal faces and evaluate its ability to classify human faces that have been intentionally generated to appear "animal-like." Our approach utilizes a classic machine learning pipeline, combining principal component analysis (PCA) for feature extraction and a support vector machine (SVM) for classification, allowing us to probe the capabilities and limitations of these established methods on a novel, cross-domain task.

The classifier performed effectively on its source domain of animal images, confirming the model was well-trained. However, when applied to AI-generated human faces, the model's performance dropped significantly, highlighting the inherent difficulty of transferring learned features across disparate domains. To investigate whether this limitation is specific to our classical pipeline, we conducted a comparative experiment with a modern deep learning model. The superior performance of the deep learning model suggests that the challenge lies primarily in the PCA-based feature extraction method. While the challenges are significant, a system capable of bridging this domain gap could enable novel tools for entertainment and digital avatar generation, which served as a key motivation for this exploratory study.

This paper is organized as follows. Section 2 details our methodology, Section 3 presents the experimental setup and results, and Section 4 concludes with a discussion of our findings and future work.

2 Proposed Method

This section details the workflow of our classification pipeline, which is illustrated in Figure 1. Our approach consists of two main stages: feature extraction using PCA and classification using an SVM.

2.1 Data Collection and Preprocessing

The model was trained and evaluated using a custom, balanced subset created from the "Animal Faces" dataset, an open-source collection sourced from Kaggle [3]. For this study, we curated a dataset consisting of three classes, which we will refer to as **cat**, **dog**, and **tiger**. The tiger images were specifically extracted

*Project Leader, Report Writer
†Data Collector, Experiment Implementation
‡Data Analyst, Code Contributor
§Presenter, Content Strategist

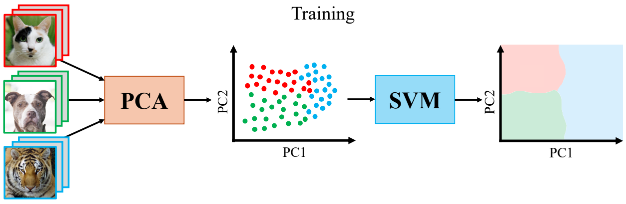
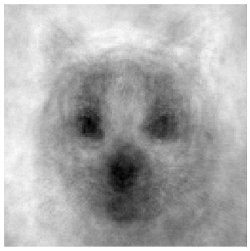
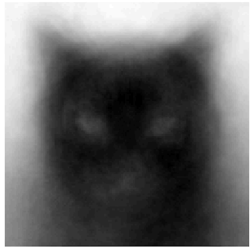


Figure 1: The training pipeline for the animal face classifier.



(a) Eigenface1



(b) Eigenface2

Figure 2: The top two principal components (Eigenfaces) extracted from the training dataset.

from the original "wildlife" category to create a more focused class. Each class comprises 450 images, resulting in a total dataset of 1,350 images.

All images underwent the same standardized preprocessing pipeline: conversion to grayscale, resizing to a uniform resolution of 128×128 pixels, flattening into a 16,384-dimensional feature vector, and normalization of pixel values to a range between 0 and 1. The complete dataset was then partitioned into a training set (70%), a validation set (15%), and a test set (15%) for model development, hyperparameter tuning, and final evaluation.

2.2 Feature Extraction: PCA

Following preprocessing, we employed PCA [1] to reduce the high-dimensional pixel data into a lower-dimensional feature space. PCA identifies a set of orthogonal axes, known as principal components, that capture the maximum variance in the training data. By projecting the data onto a subspace spanned by the top components, we can drastically reduce dimensionality while retaining the most significant feature information. In this study, we utilized the 'PCA' implementation from the scikit-learn library [4]. We reduced the dimensionality from 16,384 to 110 components, which collectively captured over 80% of the original variance in the dataset.

2.3 Classification: SVM

For the classification task, we used a Support Vector Machine (SVM) [1], a supervised learning algorithm that finds an optimal hyperplane to separate data points into different classes. Given its effectiveness in high-dimensional spaces, SVM is an appropriate choice for classifying the features extracted by PCA. We employed the 'SVC' (Support Vector Classification) module from scikit-learn. To handle non-linear relationships within the feature space, a radial basis function (RBF) kernel was selected. The key hyperparameters of the model, namely the regularization parameter C and the kernel coefficient γ, were tuned using a grid search, with the best combination selected based on performance on the validation set.

3 Experiments and Evaluation

To evaluate the capabilities and limitations of our trained classifier, we conducted two primary experiments. The first experiment assesses the model's ability to classify human faces with perceived animal-like traits, testing the hypothesis that features learned from an animal domain can be transferred to a human domain. The second experiment provides a performance baseline by comparing our classical PCA+SVM pipeline against a deep learning model. This section first details the setup and baseline validation of our PCA+SVM model, which forms the basis for these subsequent experiments.

3.1 Experimental Setup

All experiments were conducted in a Python environment, primarily utilizing the scikit-learn and OpenCV [2] libraries. Our model's configuration was based on the methods described in Section 2. The PCA model was set to reduce the feature space from 16,384 to 110 components, a number chosen to ensure that over 80% of the total statistical variance in the training data was retained. Figure 2 visualizes the top two principal components extracted from the dataset.

The subsequent SVM classifier employed a RBF kernel. Its key hyperparameters, namely the regularization parameter C and the kernel coefficient γ, were tuned using a grid search with cross-validation to maximize classification performance, resulting in optimal values of $C = 10$ and $\gamma = \texttt{scale}$[1]. To evaluate model performance, we use the following metrics. Accuracy is the ratio of correctly predicted instances to the total number of instances, serving as a general performance measure. Recall measures the proportion of actual positive instances for a given class that were correctly identified by the model. The F1-score is the harmonic mean of precision (the proportion of positive predictions that were correct) and recall, providing a single metric that

[1] In scikit-learn, the `scale` option for the γ parameter automatically sets its value to $1/(n\_features \times X.var())$, providing a data-dependent heuristic that adapts to the variance of the features.

Table 1: Classification performance on the animal test set (baseline).

Class	Precision	Recall	F1-Score
Cat	0.84	0.91	0.87
Dog	0.90	0.81	0.85
Tiger	0.90	0.91	0.91
Overall accuracy			0.88

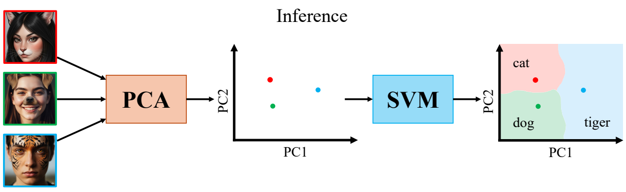


Figure 3: The inference pipeline for classifying animal-like human faces.

balances both. It is calculated as follows:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Before applying the model to the cross-domain task, we established its baseline performance on the held-out test set of animal images. As shown in Table 1, the classifier achieved a high overall accuracy of 0.88. The class-wise metrics, including precision, recall, and F1-scores, were also consistently high, confirming that the PCA+SVM model is robust and well-generalized for the task of classifying animal faces.

3.2 Experiment 1: Identification of Animal-like Human Faces

The primary objective of this experiment was to evaluate whether the PCA+SVM model, trained exclusively on animal faces, could successfully classify human faces exhibiting perceived animal-like traits. This experiment directly tests the hypothesis that the feature space learned from the animal domain is transferable and meaningful for identifying analogous subjective features in the human domain.

The method for this experiment is visualized in Figure 3. To create a test set for this cross-domain task, we generated 30 images of animal-like human faces using AI image generation services (e.g., Pollinations.ai, ChatGPT). The dataset was balanced with 10 images for each of the three target classes: **cat-like**, **dog-like**, and **tiger-like**. Ground truth labels were assigned based on the text prompts used during generation. These images were then processed through the same pipeline as the animal data and classified by the pre-trained SVM model detailed in Section 3.1.

The detailed classification performance on the 30 animal-like human faces is presented in Table 2. The overall accuracy for this cross-domain task was 0.33 (10 out of 30 images correct). The results indicate that performance was highly imbalanced across the classes. The dog-like class achieved the highest recall (0.70), correctly identifying 7 out of 10 images. In contrast, the model struggled significantly with the other classes, achieving a recall of only 0.20 for tiger-like and 0.10 for cat-like faces. An interesting result is the perfect precision (1.00) for the tiger-like class, which suggests that while the model rarely predicted a face as tiger-like, its predictions for this class were always correct.

Table 2: Classification performance on the “animal-like” human faces test set.

Class	Precision	Recall	F1-score
Cat-like	0.12	0.10	0.11
Dog-like	0.35	0.70	0.47
Tiger-like	1.00	0.20	0.33
Overall accuracy			0.33

The low overall accuracy of 0.33 confirms the significant domain gap between the animal training data and the human test images. The features learned via PCA from real animal photographs do not effectively generalize to the subtle cues that constitute an animal-like appearance in humans. The detailed metrics reveal a complex failure mode. The model is most successful at identifying dog-like faces (recall of 0.70), suggesting some overlap between the learned features of real dogs and the characteristics of these human images. Conversely, the extremely low recall for cat-like faces (0.10) indicates a severe feature mismatch. The most notable finding is the perfect precision (1.00) but very low recall (0.20) for the tiger-like class. This implies that the model has learned a very strict and specific definition for the tiger class from the training data, making it highly conservative in its predictions. This raises the critical question of whether this poor performance is a result of a fundamental, unbridgeable gap between the animal and human domains, or a limitation of the classical PCA+SVM pipeline itself. The following experiment aims to investigate this question by comparing our model with a modern deep learning architecture.

3.3 Experiment 2: Comparison with Deep Learning Models

The objective of this experiment was to provide a performance baseline for our classical PCA+SVM pipeline by comparing it against a modern deep learning architecture. This comparison aims to determine if the challenges observed in Experiment 1 are specific to our chosen methodology or represent a more fundamental difficulty in this cross-domain classification task.

For this comparison, we employed a pre-trained YOLOv8 (you only look once) model, a state-of-the-art deep convolutional neural network [5]. The YOLO model was trained on the same animal dataset as the PCA+SVM model. It was then used to classify the same set of 30 animal-like human face images used in Experiment 1.

The results of the comparison are presented in Table 3. The YOLO-based model achieved an overall accuracy of 0.63, substantially outperforming the 0.33 accuracy of the PCA+SVM model. The performance improvement was consistent across all classes, with the most significant gains observed in the cat-like (recall of 0.40 vs. 0.10) and tiger-like (recall of 0.70 vs. 0.20) categories.

Table 3: Performance Comparison of PCA+SVM vs. YOLOv8 on "animal-like" human faces.

Metric	PCA+SVM	YOLOv8
Recall		
Cat-like	0.10	0.40
Dog-like	0.70	0.80
Tiger-like	0.20	0.70
Overall accuracy	0.33	0.63

The substantially higher accuracy of the YOLO model suggests that the challenges observed in Experiment 1 are not fundamental to the cross-domain task itself, but are instead significant limitations of the PCA-based feature extraction method. The convolutional layers in the YOLO model likely learn more complex, hierarchical, and localized features (e.g., textures, shapes of specific parts like ears or eyes) that are more robust and transferable across domains than the global, variance-based features captured by PCA. While the YOLO model also did not achieve perfect accuracy, its superior ability to identify cat-like and tiger-like faces indicates that its learned feature space is more aligned with the subtle cues that constitute these human-perceived traits. This highlights the advantage of end-to-end feature learning in deep models for complex and abstract classification tasks.

4 Conclusion

In this study, we investigated the feasibility of using a classical PCA+SVM pipeline, trained exclusively on animal photographs, to classify subjective animal-like traits in human faces. While the model achieved high accuracy on the source domain of animal classification (0.88 on the test set), its performance dropped significantly to 0.33 on the cross-domain task of identifying these traits in human images. The key insight from our experiments is the confirmation of a severe domain gap. The global, variance-based features extracted by PCA from animal photographs do not effectively generalize to the subtle cues that constitute human perception of these traits. The superior performance of a YOLO-based model in a comparative experiment further suggests that this limitation is characteristic of the classical feature extraction method, and that learned, hierarchical features are better suited for such abstract problems. Future work could focus on bridging this domain gap using techniques such as data augmentation and domain adaptation (e.g., fine-tuning). A broader comparison with other machine learning architectures could also provide deeper insights into the most effective approaches for this unconventional classification challenge.

Acknowledgments

We would like to express our sincere gratitude to our supervisor, Dr. Suyong Eum, for his invaluable guidance and support throughout this project.

References

[1] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[2] G. Bradski. The opencv library. *Dr. Dobb's Journal of Software Tools*, 2000.

[3] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[4] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake VanderPlas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, 12(85):2825–2830, 2011.

[5] Dillon Reis, Jordan Kupec, Jacqueline Hong, and Ahmad Daoudi. Real-time flying object detection with YOLOv8, 2023. Preprint at `https://doi.org/10.48550/arXiv.2305.09972`.