Applied Machine Learning: Spam Detection & Face Alignment Report

Candidate Number: 279187

Abstract:

For face alignment, we down-sample 256x256 images to 96x96, augment (horizontally flip, and brightness contrast) to double the training set, and utilize two descriptors: HOG (23k dimensions) and PCA-compressed grid-SIFT. HOG achieves a **4.11 pixel mean landmark error** and places 69 % of landmarks within 0.05 inter-ocular-distance on validation, outperforming SIFT (4.76 px). These outcomes demonstrate that carefully engineered "shallow" pipelines, targeted augmentation, and rigorous validation can deliver high accuracy without heavy computational cost.

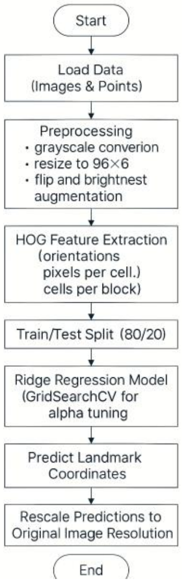
Face Alignment

Introduction:

Face alignment, the precise localization of semantic facial key points/landmarks such as eyes, nose tip, and mouth corner, is a fundamental feature in many computer vision applications, including but not limited to face recognition, expression analysis, and human-machine interaction. This section outlines the methodical design, implementation, and evaluation of a machine learning pipeline (Flowchart 2) developed for this purpose. The primary system developed and analysed employs Histogram of Oriented Gradients (HOG) features as the core image descriptor, paired with a regularized linear regression model (Ridge Regression / lecture 15) for predicting the coordinates of five facial landmarks. An alternative approach utilizing Scale-Invariant Feature Transform (SIFT) features extracted from a dense grid was also implemented and evaluated as a comparative baseline to contextualize and justify the choice of HOG. The overarching goal is to demonstrate a well-reasoned system design, precise reproducibility (random seed = 42), data preprocessing and augmentation strategies, robust feature engineering, and a critical appraisal of the chosen system's performance and inherent limitations.

Methodology:

A structured pipeline (Flowchart 2) was implemented, encompassing data preparation, image preprocessing, data augmentation, HOG feature extraction as the primary method, with SIFT as an explored alternative, and finally, training a Ridge Regression model with appropriate hyperparameter tuning via cross-validation. Reproducibility was a key consideration, addressed by using fixed random states in stochastic processes like data splitting and model initialization.


Start
Load Data (Images & Points)
Preprocessing
• grayscale converion
• resize to 96×6
• flip and brightnest augmentation
HOG Feature Extraction (orientations pixels per cell.) cells per block)
Train/Test Split (80/20)
Ridge Regression Model (GridSearchCV for alpha tuning
Predict Landmark Coordinates
Rescale Predictions to Original Image Resolution
End

Flowchart 2: Task 2 Pipeline

Data Preprocessing and Augmentation:

The dataset consisted of 2,811 training and 554 test 256x256 grayscale images with 5 associated landmark coordinates (*face_alignment_*.npz*). Following assignment guidance, images were resized to 96x96 pixels using *cv2.INTER_AREA* for computational efficiency, with landmark coordinates scaled proportionally. Deterministic augmentation was applied to improve robustness and expand the dataset: each training image produced two samples – the original **(resized)** and an augmented version (horizontally flipped with adjusted landmarks, plus a consistent brightness/contrast shift). This doubled the training set to 5,622 samples. Justification: Resizing speeds up processing; augmentation increases data diversity, improving model generalization to variations in orientation and lighting.

Feature Engineering: HOG was selected as the primary feature descriptor for its object detection and shape description efficiency. The implementation of feature extraction was implemented using *skimage.feature.hog* with parameters: *orientations=9, pixels_per_cell=(5, 5), cells_per_block=(3, 3), transform_sqrt=True, block_norm='L2-Hys' yielding a **23,409-dimensional feature vector** for each 96x96 augmented image.* This configuration of parameters (Table 5) provides a rich representation of the 96x96 input gradient patterns indicative of facial structures, with normalization providing illumination resistance.

- **Alternatively Explored: Grid-SIFT:** As a point of comparison, SIFT features (Table 5) were extracted from a *10x10 grid (kp_size=3),* resulting in 12,800 raw dimensions. These high-dimensional features were processed with *StandardScale*r and *PCA (retaining 99% variance)*, reducing them to 3,252 dimensions before feeding them into Ridge regression. (This path served as a comparative baseline.)

Descriptor	HOG
Orientations:	9
Pixels-per-cell	5, 5
Cells-per-block	3, 3
PCA	-
Dimension	23,409
Descriptor	SIFT
Grid	10x10
Kp Size	3
Raw	12,800D
PCA	99%
Post PCA	3,252

Table 1: Feature Extraction parameters

Regression Modeling Pipeline: In HOG features are well-scaled as HOG vectors are already L2-Hys normalized, thus for simplicity, we use raw HOG features, omitting StandardScaler (used for SIFT feature extraction)

Hyperparameter Optimisation: The Ridge alpha (regularization strength) was tuned/optimized for the HOG pipeline using 5-fold GridSearchCV with *neg_mean_squared_error* scoring.

- alpha grid: *{1.0, 10.0, 50.0, 100.0}*.
- Best alpha found for HOG+Ridge: 50.0.
- *(The SIFT+PCA+Ridge alternative's best alpha was 10000).*

Results:

The optimized HOG+StandardScaler+Ridge system (alpha=50.0) was evaluated on a hold-out validation set (1,124 samples, 20% of augmented data). Predicted coordinates were scaled back to the original 256x256 resolution for pixel error reporting.

- **Evaluation Metrics:** Performance was assessed using established metrics in face alignment:
    - **Mean Euclidean Error (MEE):** The average pixel distance between each predicted landmark and its corresponding ground truth, averaged over all five landmarks and all images in the validation set. (Lower values indicate better performance)
    - **Per-Landmark Mean Error:** The MEE calculated individually for each of the five landmark types (left eye, right eye, nose tip, left mouth, right mouth).
    - **Mean Inter-Ocular Distance (IOD) Normalized Error:** The MEE for each landmark normalized by the ground truth IOD (the distance between the centers of the two eyes) for that image. This provides a scale-invariant error measure.
    - **Percentage of Landmarks within IOD Thresholds:** The proportion of landmarks whose IOD-normalized error is less than 0.05 and 0.10. Higher percentages indicate better precision for a larger number of predictions.

- **Quantitative Performance:** The HOG-based system achieved accurate results on the validation set, confirming its suitability.

Metric	HOG + Ridge (alpha = 50)	SIFT + Ridge (alpha = 10,000)
Mean Error (px)	**4.11**	**4.76**
L-Eye Error (px)	2.91	3.21
R-Eye Error(px)	2.89	3.32
Nose Error (px)	5.30	6.18
L-Mouth Error(px)	4.63	5.52
R-Mouth Error(px)	4.81	5.55
Mean IOD Norm Error	0.0427	0.0495
% Landmarks < 0.05 IOD	69.1%	61.2%
% Landmarks < 0.10 IOD	94.3%	91.4%

Table 2: Hold-Out Validation Performance Comparison for SIFT and HOG

As shown in Table 6, the primary HOG+Ridge system achieved an overall Mean Euclidean Error of 4.11 pixels and a Mean IOD Normalized Error of 0.04275 on the validation set. Notably, 69.1% of landmarks were localized within an IOD-normalized error of 0.05, and 94.3% within 0.10. The per-landmark errors indicate strong performance for eye localization and reasonable accuracy for the nose and mouth corners. These results are superior to those achieved by the SIFT+PCA+Ridge pipeline (MEE 4.76 pixels, 61.2% < 0.05 IOD), reinforcing the selection of HOG as the more effective feature descriptor for this task configuration.

- **Cumulative Error Distribution (CED) Curve:** The CED curve provides a comprehensive visual summary of the error distribution for the HOG+Ridge system across the validation landmarks.

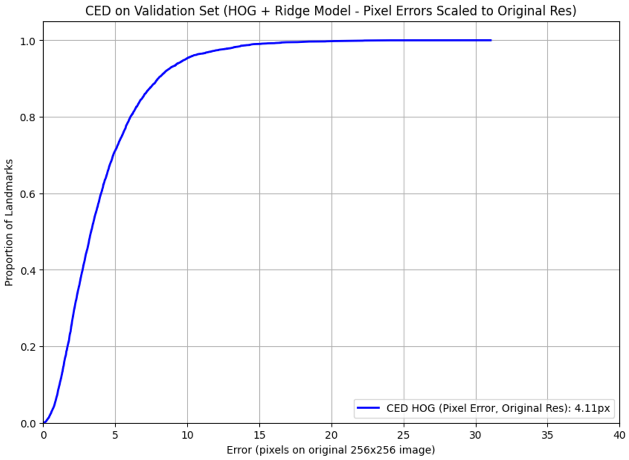


Figure 6: Cumulative Error Distribution (CED) Curve for HOG+Ridge System

Figure 6 illustrates that the HOG+Ridge system localizes a substantial fraction of landmarks with low errors; Approximately 94.3% of landmarks are predicted with an error of less than 10 pixels and 69.1% with an error of less than 5 pixels on the 256 × 256 scale, confirming HOG's high accuracy across various error thresholds and landmarks.

Failure Cases and Critical Analysis:

A critical aspect of a face alignment model is not only to build a functional system but also to analyze its performance, including its limitations and typical failure modes. This demonstrates an understanding of the chosen techniques and their applicability.

Analysis of HOG+Ridge Failure Modes: While the HOG+Ridge system performed well on average, inspection of predictions on the validation set revealed specific scenarios where accuracy degraded:

- **Atypical Expressions/Poses:** Landmark predictions, particularly for mouth corners, were less accurate on faces with substantial deviations from neutral expressions (e.g., wide grins) or significant out-of-plane rotation. These variations likely alter local HOG patterns substantially compared to the training data norm.
- **Illumination/Occlusion Issues:** Extreme lighting conditions (Figure 8) (hard shadows, saturation) or minor occlusions (e.g., hair across an eye, glasses / Figure 7) sometimes degraded HOG feature quality in affected regions, impacting prediction precision.

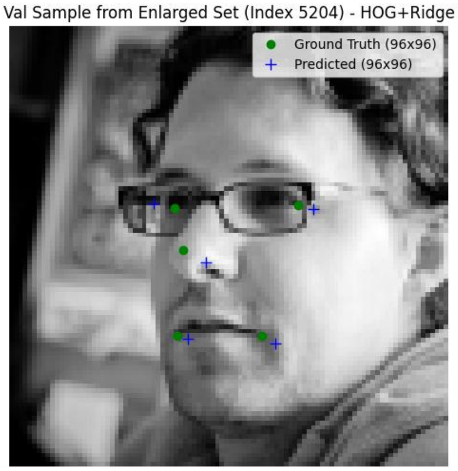


Figure 7: Misallocation of Left and Right Eye (blue crosses) compared to ground truth (green circle) because of glasses in low image quality (scaled 96x96)

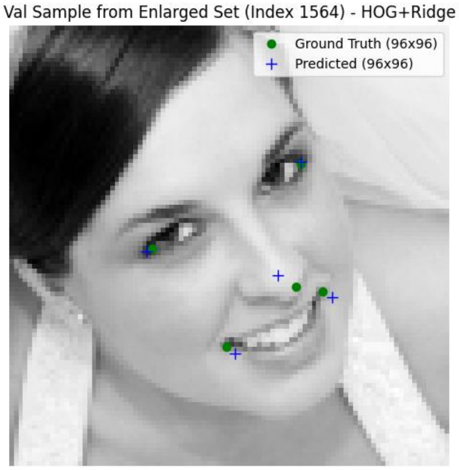


Figure 8: Inaccurate nose tip prediction under harsh side lighting.

Biases and limitations: Accuracy drops for poses, lighting or demographics under-represented in the 2,811-image training set. Rotation-sensitive, single-scale HOG offers only local edges, and Ridge captures linear links, so large appearance changes or non-linear patterns remain hard. Detail is lost in 96 × 96 downsizing, and simple flip/brightness augmentation omits occlusions and elastic deformations, further curbing robustness.

Critical Conclusion: HOG+Ridge (alpha=50) pipeline provides an effective baseline for face alignment using classical methods, achieving good average accuracy (MEE 4.11 pixels) on the validation set and outperforming the SIFT-based alternative. While adhering to Occam's Razor by starting with simpler, interpretable models, the critical analysis reveals limitations in handling significant appearance variations due to dataset constraints, HOG feature properties, and model linearity (lecture 14). While demonstrating the power of well-engineered fundamental pipelines, achieving state-of-the-art robustness would likely require addressing these aspects, particularly through more sophisticated data augmentation or models capable of capturing non-linearities.