Digital Signal and Image Management Project

Julius Maliwat 864520

Overview

Environmental Sound Classification

Classifying 50 environmental sound categories (ESC-50) by comparing traditional ML (Random Forest, SVM) with deep learning models (YamNet, AST)

Low-Light Object Detection

Detecting objects in low-light images (ExDark) and assessing the impact of CLAHE preprocessing

Vehicle Image Retrieval

Retrieving similar vehicle images (Cars196) and evaluating optimization techniques for improved retrieval.

Task 1: Environmental Sound Classification

Dataset Overview

ESC-50 contains 2000 labeled audio clips, divided into 50 classes spanning 5 categories: animals, nature, human activities, urban sounds, and musical instruments

Objective

Classify environmental sounds by comparing machine learning (Random Forest, SVM) and deep learning models (YamNet, AST)

Challenges

Variability within some classes, background noise, and similarities between sounds (e.g., rain vs. sea waves) make classification challenging

Methods overview

Traditional Machine Learning

Handcrafted features (MFCCs, ZCR, Spectral Centroid) were extracted to capture audio characteristics. Multiple classifiers (Logistic Regression, SVM, Random Forest) were evaluated, and SVM achieved the best accuracy.

Pretrained CNN-Based Feature Extractor + FCNN

YamNet (pretrained on AudioSet) extracts 1024-D embeddings from waveforms, pooled over time to obtain a fixed-length representation.
FCNN: 4 layers (512, 256, 128, 64 neurons), ReLU activations, Batch Normalization, Dropout.

Pretrained Transformer-Based Feature Extractor + FCNN

AST (pretrained on AudioSet) extracts embeddings from spectrograms, pooled over time to obtain a fixed-length representation.
FCNN: 4 layers (512, 256, 128, 64 neurons), ReLU activations, Batch Normalization, Dropout.
The same FCNN architecture as YAMNet is used to ensure a fair comparison of feature extraction methods (CNN vs Transformer).

Results

Model	Accuracy (%)
Traditional ML (SVM)	46.3
YamNet + FCNN	78.7
AST + FCNN	**94.2**

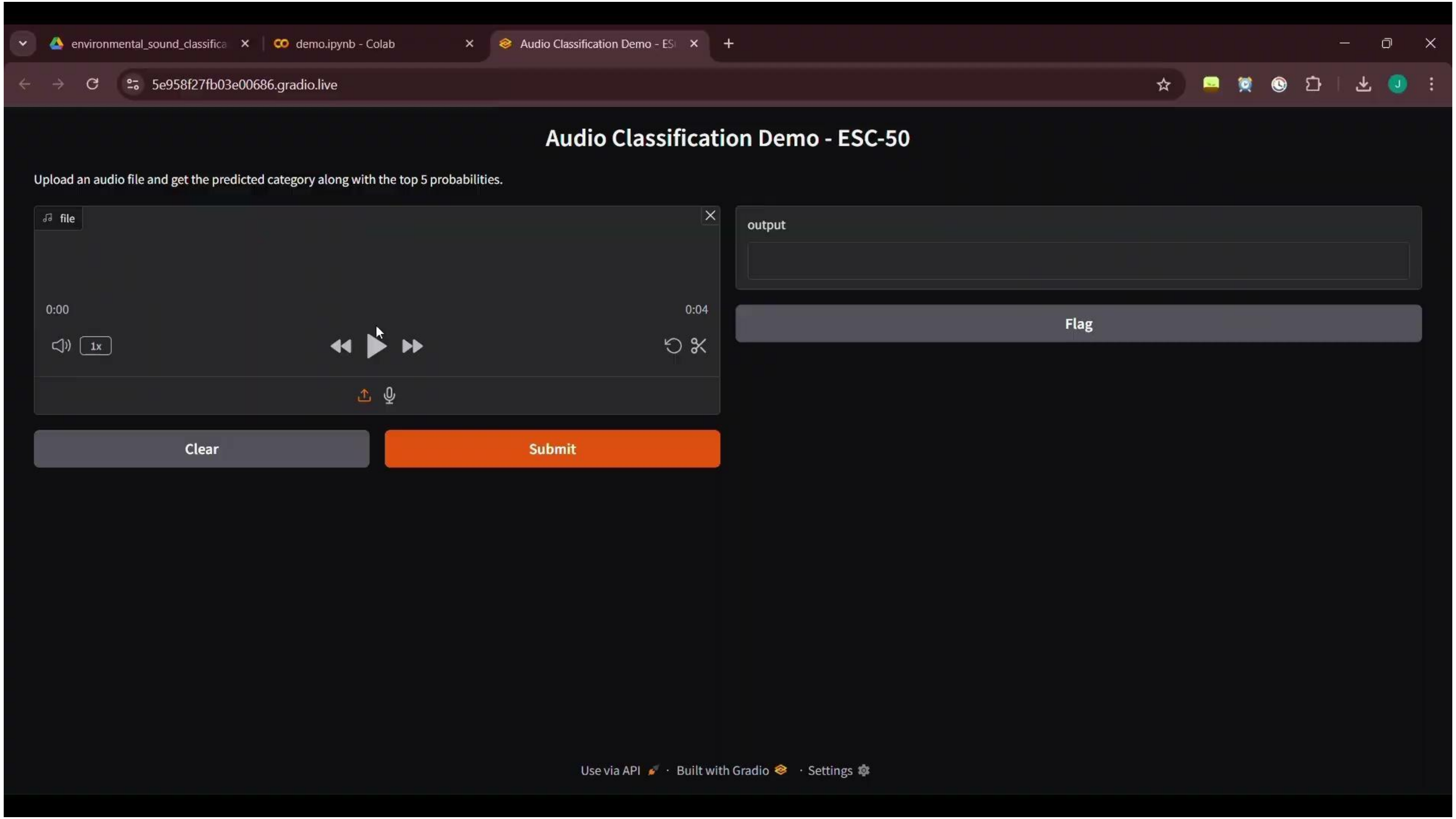
Results obtained using 5-Fold Cross-Validation.
Reported values represent the mean accuracy.

Best classified sounds:

Crow, Dog, Frog, Rooster, Toilet Flush (F1-score = 1.00).

Most challenging sounds:

Helicopter (F1 = 0.76), Washing Machine (F1 = 0.80), Engine (F1 = 0.80).

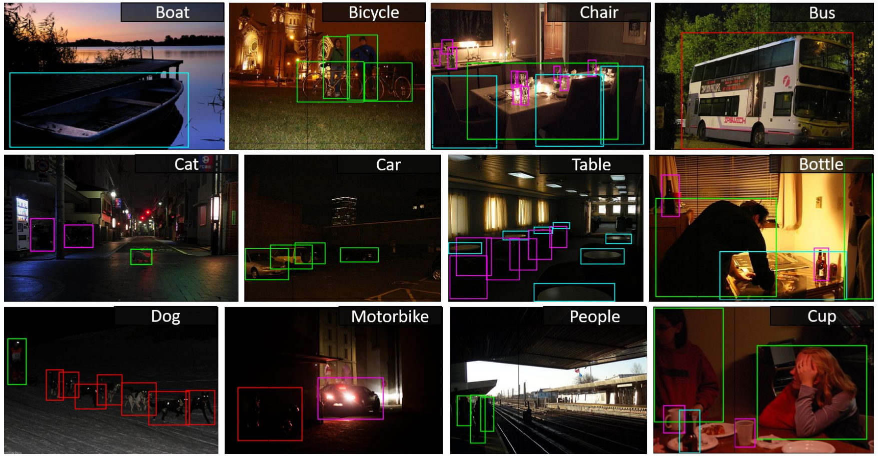

environmental_sound_classifica
demo.ipynb - Colab
Audio Classification Demo - ES
5e958f27fb03e00686.gradio.live
Audio Classification Demo - ESC-50
Upload an audio file and get the predicted category along with the top 5 probabilities.
file
0:00
0:04
1x
output
Flag
Clear
Submit
Use via API · Built with Gradio · Settings

[Demo]

Task 2: Low Light Object Detection

Dataset Overview

Collection of low-light images across 12 object categories (e.g., bicycle, bus, car, dog, people), featuring diverse lighting conditions that challenge traditional object detection models



Objective

Evaluate object detection in low-light images and assess the impact of CLAHE preprocessing on detection performance

Challenges

Low contrast, high noise, and varying illumination conditions challenge object detection models.

Methods

CLAHE-Based Preprocessing

Adaptive CLAHE applied based on mean luminance levels, dynamically adjusting clip limit and tile grid size

Object Detection Model

YOLOv8 Nano, pretrained on COCO, fine-tuned on ExDark.

Same training configuration for both raw and CLAHE-enhanced images.

First 10 layers frozen to retain pretrained feature extraction. Trained for 20 epochs on ExDark.

Validation Results

Approach	mAP@0.5	mAP@0.5-0.9	Precision	Recall
Baseline	63.37	39.78	67.8	57.5
CLAHE Enhanced	**64.5**	**40.27**	**68.4**	**58.6**

CLAHE preprocessing showed a slight improvement in validation performance. This configuration was selected for final testing

Test Results

Approach	mAP@0.5	mAP@0.5-0.9	Precision	Recall
CLAHE Enhanced	61.1	37.8	68.8	54.4

Per-Class Performance (mAP@0.5)

Object Class	mAP@0.5
Bus	**77.6**
Bicycle	**75.6**
People	69.8
Car	68.7
Dog	66.9
Boat	62.7
Bottle	61.3
Cup	57.6
Chair	54.2
Cat	49.2
Motorbike	49.0
Table	40.0

Best Performing Class: Bus

Bus performed the best with a mAP@0.5 of 77.6, likely due to its distinctive shape and size.

Most Challenging Class: Table

Table had the lowest mAP@0.5 (40.0), possibly due to occlusions and variations in lighting

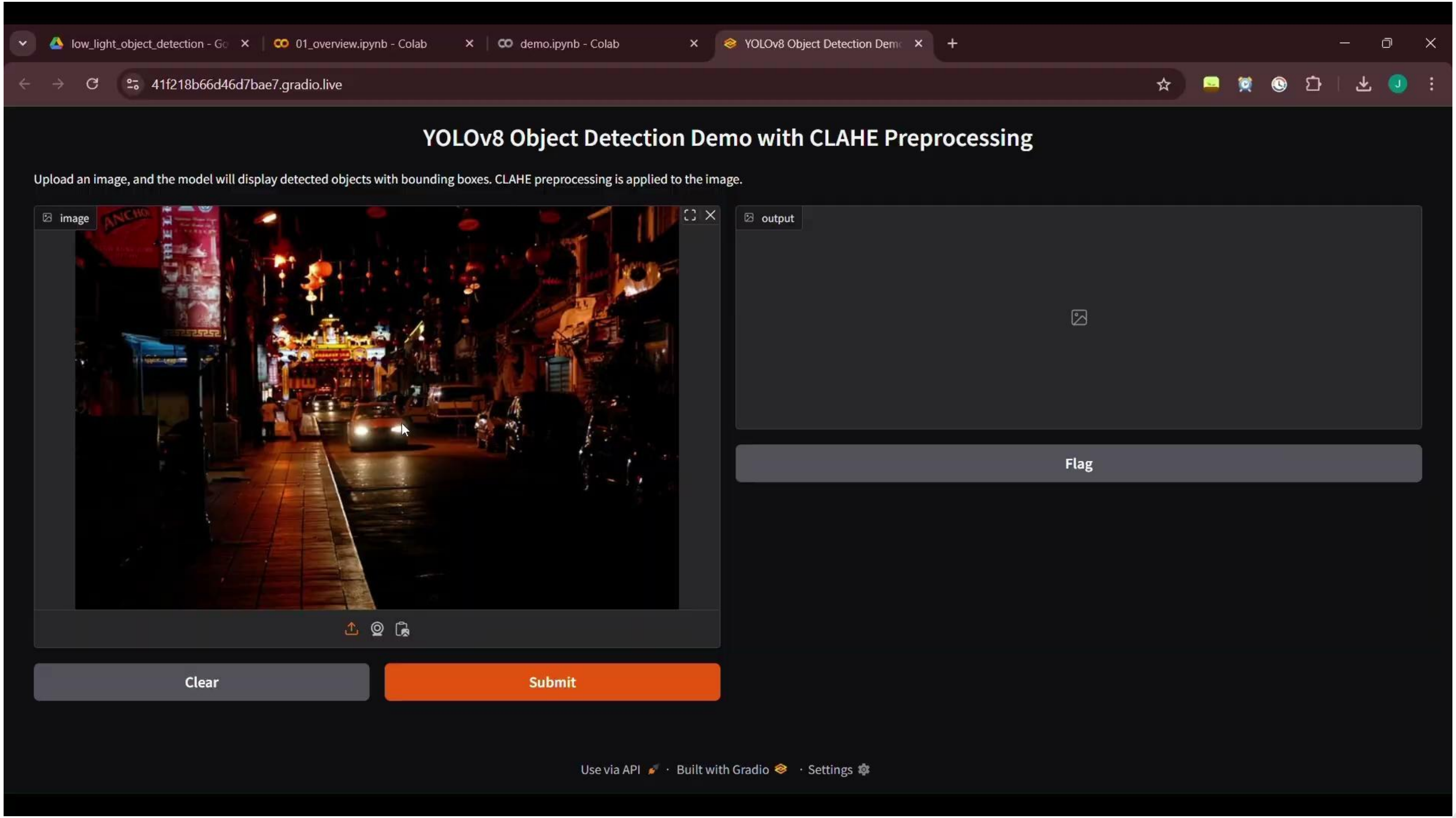
Test Results

Light Condition	**mAP@0.5**
Screen	**75.73**
Twilight	68.04
Single	64.17
Ambient	63.06
Window	61.06
Object	59.96
Strong	56.42
Shadow	55.81
Weak	53.17
Low	49.26

Detection performance significantly drops under low-light conditions, while screen lighting yields the best results.

Light Condition: **Low**

Light Condition: **Screen**



low_light_object_detection - Go
01_overview.ipynb - Colab
demo.ipynb - Colab
YOLOv8 Object Detection Dem
41f218b66d46d7bae7.gradio.live
YOLOv8 Object Detection Demo with CLAHE Preprocessing
Upload an image, and the model will display detected objects with bounding boxes. CLAHE preprocessing is applied to the image.
image
output
Flag
Clear
Submit
Use via API · Built with Gradio · Settings

[Link to Demo]

Task 3: Vehicle Image Retrieval

Dataset Overview

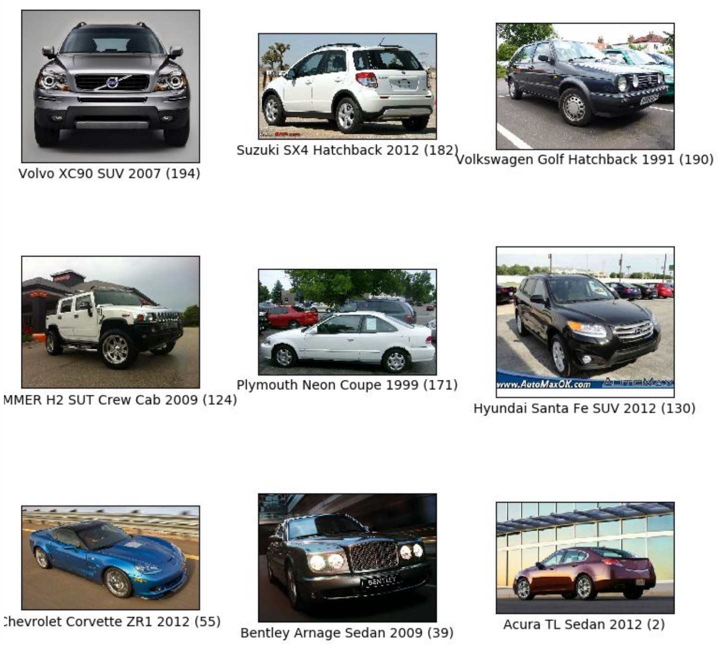
The Cars196 dataset consists of 16,185 images from 196 car classes, split into 8,144 training and 8,041 test (query) images.

Objective

Optimize vehicle image retrieval by comparing Proxy Loss and Center Contrastive Loss for learning discriminative feature representations.

Challenges

High intra-class variance (same model, different angles and lighting) and low inter-class variance (visually similar but different models) challenge retrieval.



Methods

Feature Extraction Model

EfficientNetB0 (fully fine-tuned) → GAP → L2 Normalization → 128-D Feature Embeddings

Proxy Loss

Learns class-level feature representatives (proxies) to optimize retrieval.

Center Contrastive Loss

Encourages embeddings to cluster around class centers while maximizing inter-class separation.

Training setup

The same model was fine-tuned on Cars196 for 20 epochs, testing Proxy Loss and Center Contrastive Loss separately to evaluate their impact on retrieval performance

Results

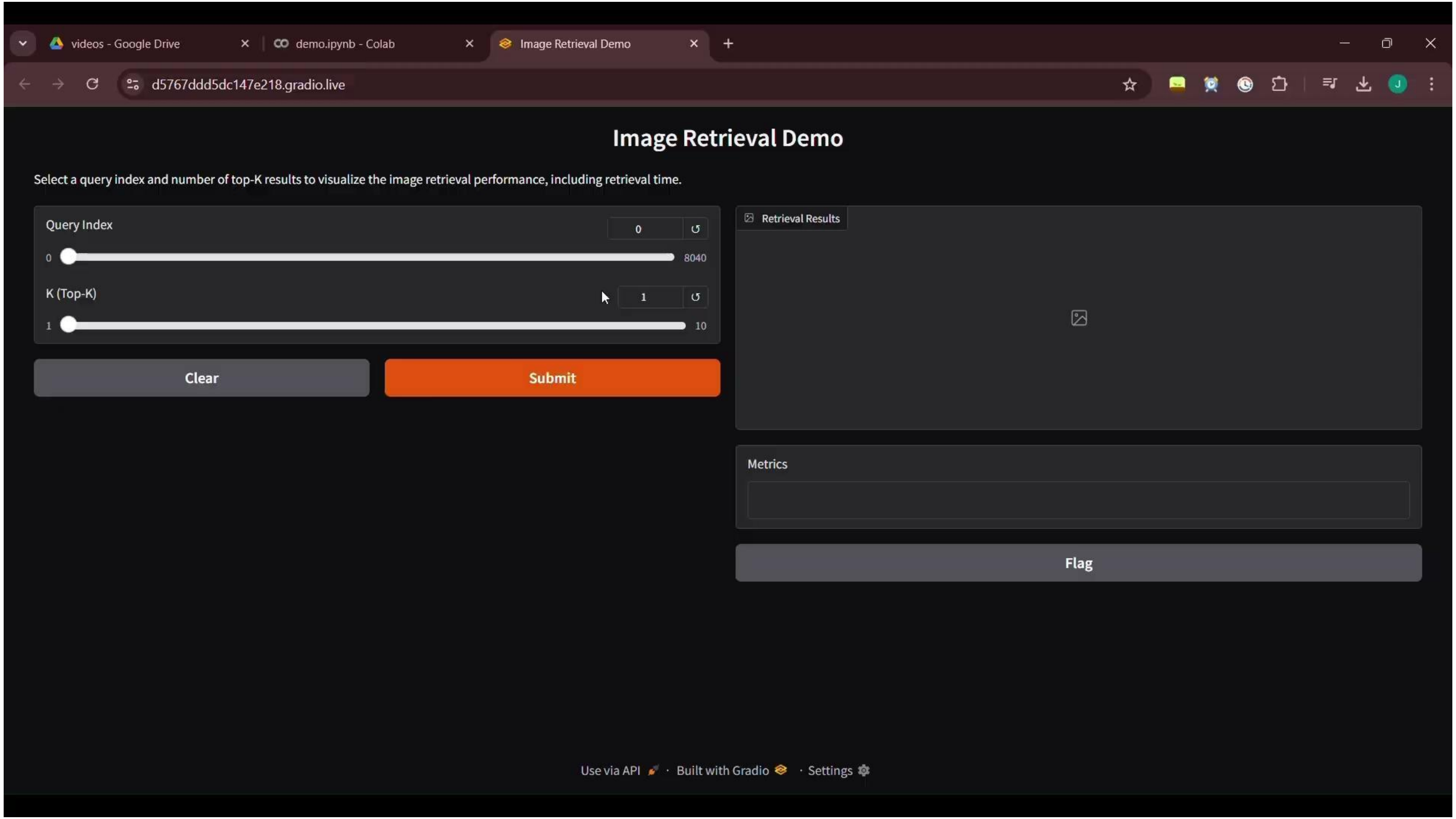
Loss Function	Recall@1	Recall@5
Proxy Loss	**67.07**	75.18
Center Contrastive Loss	64.08	**78.26**

Proxy Loss performs better for immediate retrieval (Recall@1), ensuring the most relevant match is ranked first.

However, Center Contrastive Loss slightly improves broader retrieval (Recall@5), retrieving more relevant images in the top-5 results.

Best: Jeep Wrangler SUV 2012 (Recall@1 = 97.62%)

Worst: Chevrolet Express Cargo Van 2007 (Recall@1 = 17.2%)


videos - Google Drive
demo.ipynb - Colab
Image Retrieval Demo
d5767ddd5dc147e218.gradio.live
Image Retrieval Demo
Select a query index and number of top-K results to visualize the image retrieval performance, including retrieval time.
Query Index
0
0
8040
K (Top-K)
1
1
10
Clear
Submit
Retrieval Results
Metrics
Flag
Use via API · Built with Gradio · Settings

[Link To Demo](#)

Future Work

Environmental Sound Classification

Exploring self-supervised learning (e.g., EAT) for audio representation learning could improve classification performance beyond AST

Low-Light Object Detection

Adaptive preprocessing based on lighting metadata could improve detection in extreme low-light conditions.

Vehicle Image Retrieval

Exploring stronger backbones (e.g., EfficientNetV2, ViTs) could improve feature extraction and retrieval performance

Thanks!

Do you have any questions?