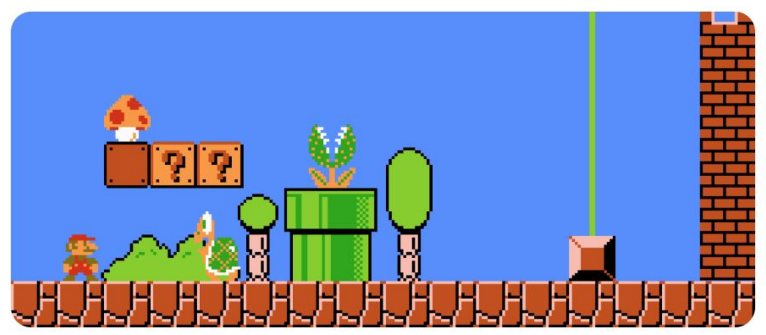
Smelling the Game: Multi-Label and Proportional Scent Prediction from Visual Scenes

Haoran Dong, Yuteng Chen, Yingyi Zhu

Github:https://github.com/QIQIZHANG852/CSCI1470_Final_Project

Introduction



In the early days of gaming, graphics were simple — often limited to just 8-bit pixel art, as seen in classics like *Super Mario*. Today, thanks to advancements in AI-generated graphics and cutting-edge rendering technologies, video games have achieved breathtaking levels of realism and detail, delivering immersive visual experiences that were once unimaginable.

This dramatic leap in visual fidelity led us, a group of avid gamers, to ask an important question:

What other senses could be enhanced to make games even more immersive?

One natural next step is the sense of smell — an often overlooked but powerful component of human experience. In this project, we explore how to bring scents into gaming by leveraging deep learning (DL) and computer vision (CV) models. Specifically, we aim to predict the scents associated with visual game scenes, creating a foundation for enhancing player immersion and interaction in future gaming experiences.

Data

Raw Source: Started from an open-source collection of ~50 000 game screenshots on Hugging Face.

Quality Filtering:

- **Semantic Relevance**: Used a pre-trained CLIP model with custom scene prompts to score and rank images by how well they matched rich environmental contexts.
- **Visual Clarity**: Applied a blur-detection network to discard any images lacking sharp detail.

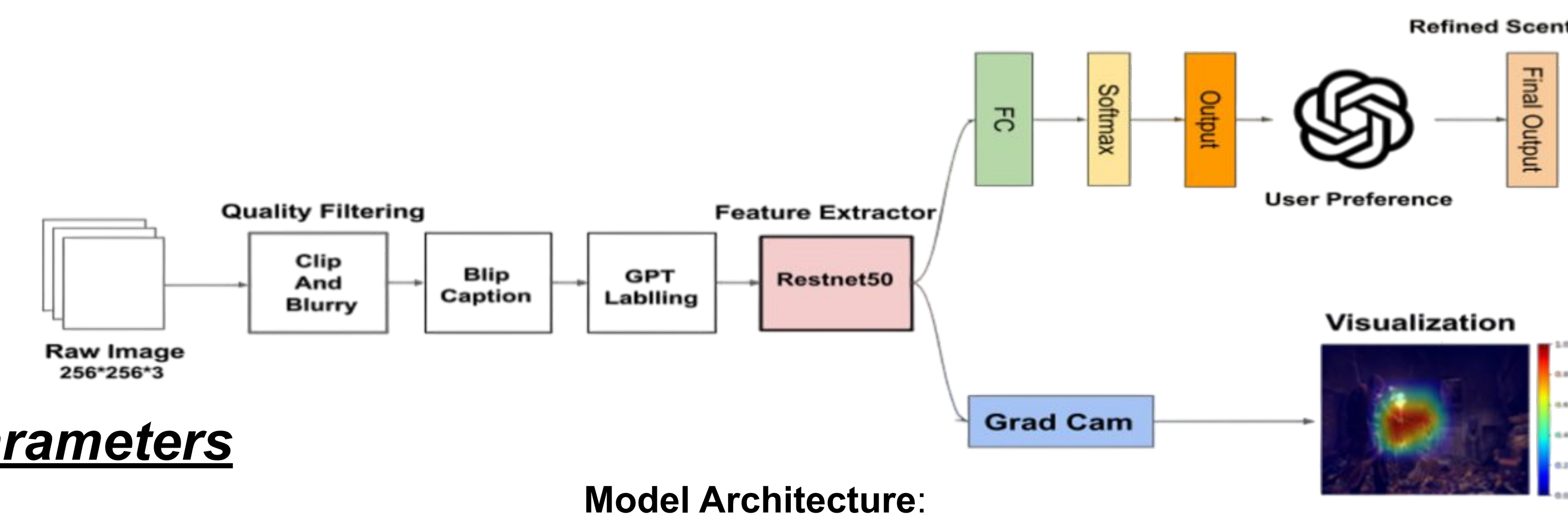
Final Selection: 2 678 high-quality, context-rich images that pass both semantic and sharpness filters.

Automated Annotation:

- **Captioning**: Employed BLIP to generate concise scene descriptions for each image.
- **Scent Labeling**: Leveraged GPT-3.5 in a zero-shot workflow, mapping captions to 21 predefined scent categories and producing normalized distributions.

Methodology

Pipeline



Hyperparameters

Batch_size	32
Epoch	20
Learning_rate	1e-5(Adam)
Activation	Relu

Model Architecture:

1. **Feature Extraction**: ResNet-50 backbone processes each 224×224 RGB frame.
2. **Classification Head**: Two fully connected layers (256→21) + softmax to output a probability distribution over scent labels.
3. **User Preference Tuning**: Optional GPT-3.5 adjustment layer refines the raw scent profile based on natural-language user feedback

Training Setup:

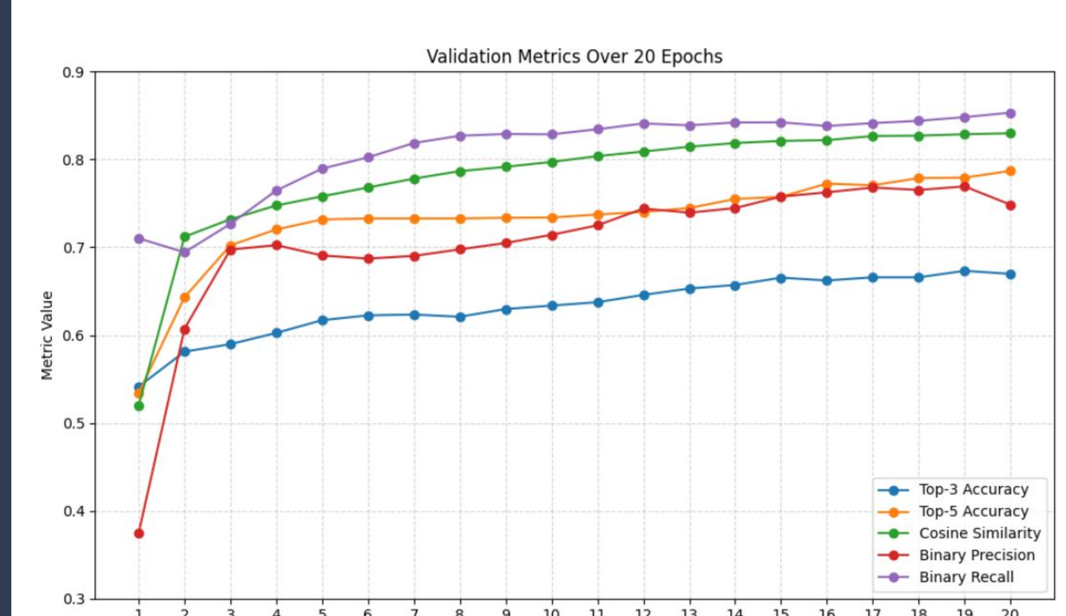
- **Data Split**: 80% train / 20% validation (fixed random seed for reproducibility).
- **Loss & Regularization**: KL divergence on soft targets + L1 output penalty to encourage sparse, decisive scent predictions.
- **Optimizer**: Adam, learning rate 1e-5, batch size 32, trained for 20 epochs.

Interpretability: Applied Grad-CAM on the final ResNet layer to visualize which image regions most strongly influenced each scent label, aiding both debugging and user trust.

Results

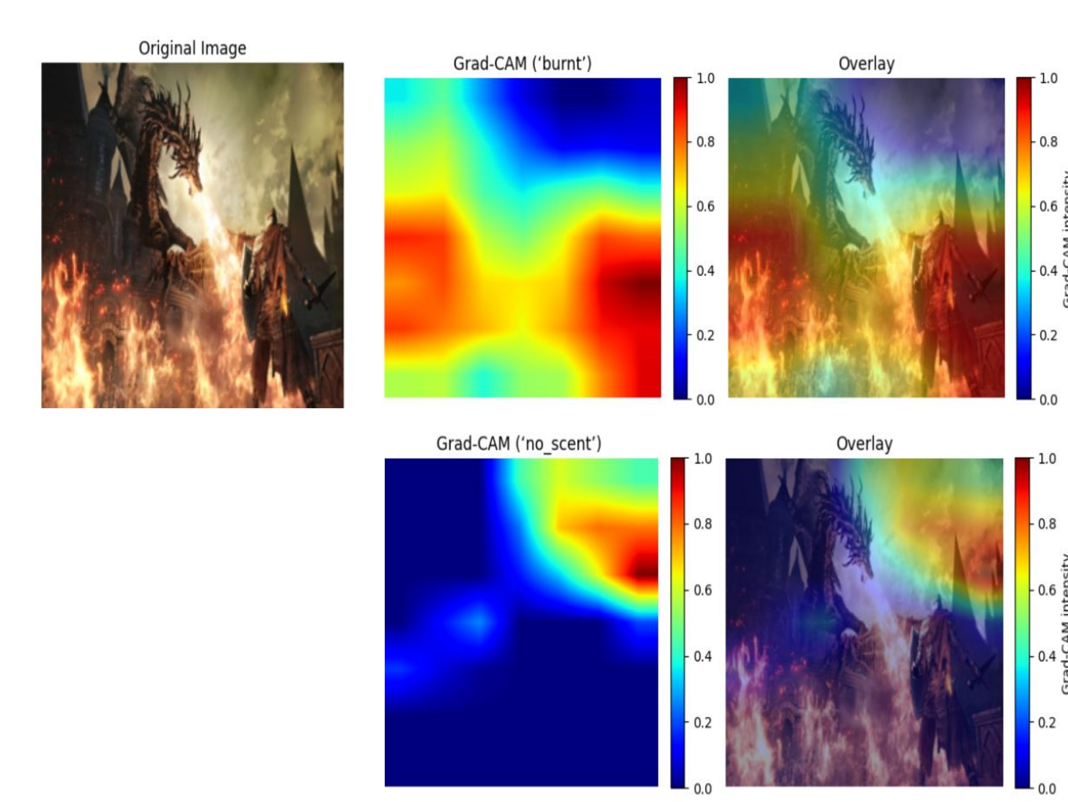
Evaluation Metrics:

- **Top-k Coverage**: Fraction of top-k true scents captured within the top-k model predictions.
- **Binary Precision & Recall**: Thresholded presence/absence accuracy per sample.
- **Mean Absolute Error (MAE)**: Average deviation across all scent proportions.
- **Cosine Similarity**: Alignment between predicted and ground-truth scent vectors.



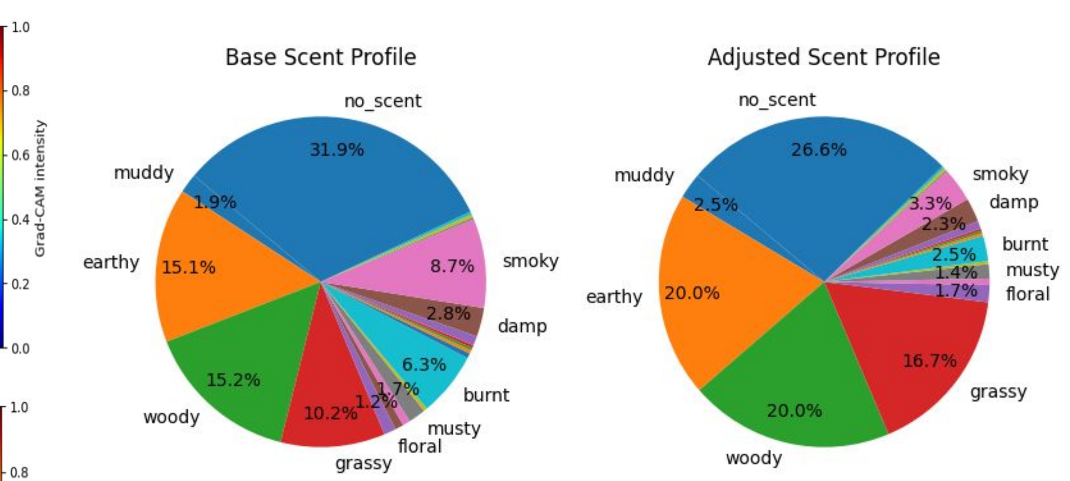
Grad-CAM Visualization:

Grad-CAM is a trick that shows you where your model is "looking" when it makes a prediction. It grabs the gradients flowing into the last convolutional layer, turns them into a heatmap, and overlays that on the original image.



User Preference Adjustment:

Example input:
"Avoid pungent smells. Prefer natural ones."



Discussion

Future Work:

1. Grow and rebalance the dataset; double rare-scent samples via targeted augmentation (color-jitter, CutMix) to lift precision
2. Replace the fixed ResNet backbone with other backbones, enabling global context and better multi-label separation
3. Tweaking hyperparameters for our ScentPredictors model

Reference:

1. Dataset: Taesiri (2024). GameplayCaptions-GPT-4V. Hugging Face. https://huggingface.co/datasets/taesiri/GameplayCaptions-GPT-4V