Abstract Submission

Members : Utsa Ghosh - 23CS10090 & Tanmay Amritkar - 23CS30066

Visual Curiosity Engine: Predicting Human Question-Provoking Regions in Images

Problem Statement : Most vision models focus on saliency prediction, i.e., estimating where humans will look in an image. Saliency is primarily driven by contrast, motion, or visual distinctiveness. However, curiosity is different from attention: people do not just look at what is visually salient, but also at what is ambiguous, novel, anomalous, or contextually surprising. The project addresses the problem of identifying regions in an image that are most likely to provoke human curiosity and questions. This project introduces a new task: predicting curiosity-inducing regions in images, i.e., the parts of an image most likely to provoke a human question. Unlike saliency maps, which capture *"where the eye goes"*, curiosity maps capture *"where the mind wonders."*

Motivation : The ability to predict curiosity-inducing regions has practical applications in marketing, where systems can automatically identify and highlight unusual or engaging product features to capture consumer interest ; in interactive assistants, which can highlight intriguing elements; and in human-AI interfaces, enabling machines to ask clarifying questions about ambiguous content. This will make AI systems more intuitive and pedagogically effective.

Challenges : Curiosity is subjective and not always aligned with visual saliency (e.g., a small but odd object can trigger more questions than a large colorful region). It requires modeling contextual mismatches, not just local distinctiveness. The most challenging aspect is that we could not find any standardized dataset for curiosity prediction.

Data Requirement : We plan to collect 200–400 images from COCO, Open Images, and VQA datasets (we can later increase the dataset to more depending on available computational power). Each image will be annotated, marking question-provoking regions with bounding boxes, assigning curiosity scores (0–5), categorizing question types (why/what/how/anomaly), and optionally providing example questions. Annotations will be aggregated into consensus heatmaps and region rankings.

Techniques/Algorithms : We are hoping to use a ResNet-50 or Vision Transformer backbone with a U-Net-style decoder for pixel-wise curiosity heatmap prediction. Region proposals from pretrained object detectors (DETR, Faster R-CNN) will be ranked using RoI-pooled features and multi-layer perceptrons. Multi-task learning will combine focal loss for heatmap prediction and pairwise hinge loss for ranking, with an optional transformer decoder for question generation..

Evaluation : We will evaluate the model using automated metrics such as Pearson correlation between predicted and human heatmaps, NDCG@k for ranking quality, and top-1 overlap accuracy. Additionally, blind A/B human studies will be conducted to compare our model with saliency-based and random baselines.

Impact : This project will establish curiosity prediction as a novel computer vision task, creating the first dataset and baseline models. The impact extends to adaptive questioning in education, more intuitive human-AI collaboration, and advancing AI systems that better align with human cognitive processes by bridging perception, cognition, and creativity.