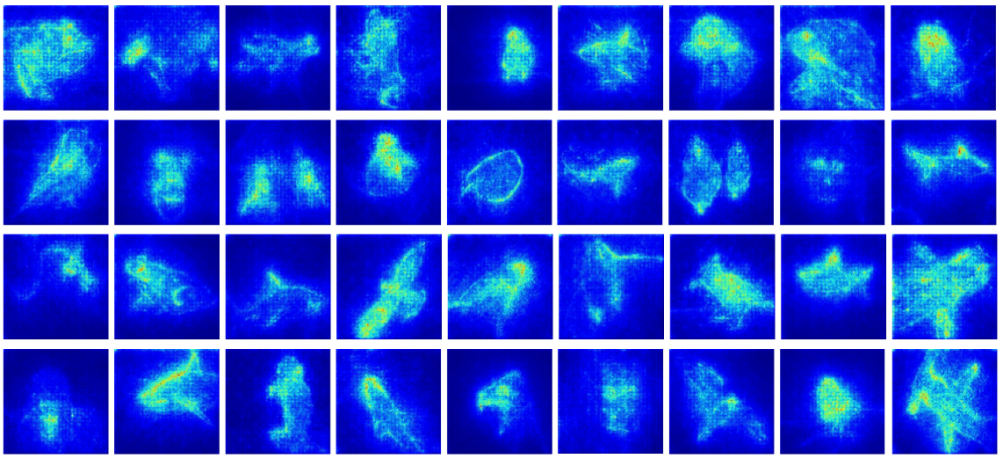
Visual Attention Prediction Using Deep Learning

Kipngeno Koech - Spring 2025 - 18-799-RW Applied Computer Vision Course

Overview

The **Visual Attention Prediction** project aims to develop a deep learning model that can predict human visual attention in images by generating **saliency maps**. These maps highlight the most visually significant regions of an image, mimicking human gaze behavior. This technology has potential applications in **human-computer interaction (HCI), assistive technologies, medical imaging, autonomous navigation, and advertising**. By leveraging convolutional neural networks (CNNs), the project seeks to improve existing methods for saliency prediction, making them more accurate and efficient.



Objective

To design and train a **deep learning model** that predicts visual attention regions in images using **CNN-based architectures**. The model will be trained on benchmark saliency datasets and evaluated using standard saliency metrics to ensure high performance.

Beyond prediction, the system will integrate with **large language models (LLMs)** to provide **contextual analysis and insights** based on the identified visual attention regions. This will enable real-time **audio or text-based feedback**, offering users **descriptive, analytical, or assistive information** about key areas of interest in an image. Such a capability can enhance **accessibility for visually impaired users, improve human-computer interaction, and support applications in education, surveillance, and medical diagnostics**.

Key Features

- **Deep Learning-Based Prediction:** The model will use CNNs (e.g., **UNet, DeepGaze II, or SalGAN**) to predict human attention maps.
- **Training on Large Saliency Datasets:** The project will utilize datasets such as

SALICON, MIT300, or CAT2000 to ensure robust learning.
- **Evaluation with Standard Metrics:** Performance will be measured using **AUC-Judd, NSS, CC, and KL divergence**.
- **Potential Real-World Applications:** The model can be applied in **HCI, marketing analytics, BCI systems, medical imaging, and autonomous navigation**.

Expected Benefits

- **Improved Visual Interaction:** Enhances AI systems' ability to process images like humans, improving user experience in HCI applications.
- **Efficiency in Decision-Making:** Can be used in **robotics, security, and medical imaging** for fast, automated decision-making.
- **Scalability:** Can be integrated into various industries requiring visual attention prediction.

Conclusion

The **Visual Attention Prediction** project represents a step toward bridging the gap between human perception and artificial vision. By leveraging deep learning, this system will enable machines to **prioritize visual information**, leading to improved decision-making in AI-powered applications. This innovation aligns with the increasing demand for intelligent, **human-like visual processing** in modern computing systems.