VISIONFUSION - A MULTI-MODAL IMAGE ANALYSIS PLATFORM

ABSTRACT

Overview:

VisionFusion is a sophisticated image analysis platform that integrates advanced computer vision and AI technologies to derive meaningful insights from visual data. Designed to go beyond mere image viewing, VisionFusion offers an extensive suite of analysis tools for a variety of applications, including object detection, segmentation, and text extraction. It can also act as a pipeline for the same task if user provides a pre-trained model on any class of image dataset.

Key Features:

- **Object Detection:** Utilizes state-of-the-art models to precisely identify and localize objects within images, producing accurate bounding boxes.
- **Segmentation:** Segregates identified objects by generating precise object masks, enhancing the accuracy of further image processing.
- **Text Extraction:** Implements advanced Optical Character Recognition (OCR) to extract textual data from images, including documents and signage.

Target Audience:

VisionFusion caters to a broad spectrum of users, including:

- **Medical Practitioners:** Derive enhanced insights from medical images to aid in disease diagnosis, such as glioblastoma.
- **Data Analysts:** Extract and analyze crucial data from environmental images, text, or microscopic imagery.
- **Artists and Designers:** Dissect artwork to uncover dominant colors, shapes, and hidden elements.

Technical Architecture:

VisionFusion integrates a variety of cutting-edge AI models and algorithms, including:

- **YOLO (You Only Look Once):** Facilitates fast and accurate object detection with real-time processing.

- **DETR (Detection Transformer):** Provides high-performance object detection and segmentation for complex scenarios.
- **OCR Engine:** Powers the text extraction feature to efficiently handle diverse types of visual data.
- **Summarization Engine:** Leverages advanced language models like LangChain and Gemini to generate concise analysis summaries.

Contributions:

- **Anubhav Mazumder (22051145):**
    - *Project Lead:* Directed conceptualization, development, and strategic oversight.
    - *Model Optimization:* Led the integration and tuning of YOLO and DETR models to balance performance and accuracy.
    - *UI/UX Designer:* Designed an intuitive interface to streamline user interaction.
- **Debjit Mandal (22051069):**
    - *Text Extraction:* Spearheaded the implementation of the OCR engine for accurate text recognition.
    - *Summarization Feature:* Developed the summarization system, using language models to create insightful reports.
    - *Documentation:* Managed project documentation and communication.

Conclusion:

VisionFusion bridges the gap between raw visual data and actionable insights through a powerful combination of AI and user-friendly design. This platform empowers diverse industries to extract value from images and opens up new possibilities for analysis and discovery.

Debjit Mandal (22051069)

Anubhav Mazumder (22051145)