Research Statement

Nahyeon Kim
November 26, 2024

My research interests lie in machine learning and computer vision, with a particular focus on diffusion models, Neural Radiance Fields (NeRF), and spatial computing. Throughout my research journey, I have contributed to innovative projects such as developing Visual Positioning Systems for VR services, integrating NeRF into development frameworks, and synthesizing high-quality images using denoising diffusion probabilistic models. Leveraging this diverse experience, my primary goal is to advance AI technologies and apply them to real-world challenges, particularly in immersive and interactive environments.

Recent Researches

Visual Positioning System (VPS) for VR/AR service To enable reality-based VR/AR services, such as the example illustrated in Figure 1, my research focused on determining user position coordinates from video captured by a device camera. I developed a Visual Positioning System (VPS) capable of predicting a user's position based on images taken with a mobile device. For the VR service, we constructed a database of the target environment using a LiDAR scanner. When a query image is sent by the user, the system identifies the most similar scene in the database and calculates the user's position matrix by applying keypoint detection, matching, and the Perspective-n-Point (PnP) algorithm.

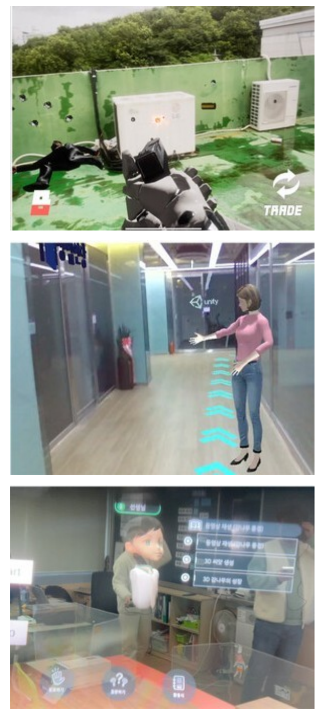

Figure 1: Actual example of VPS application on virtual education service

Optimal Image Selector for NeRF Training In Neural Radiance Field (NeRF) model training, achieving sophisticated scene rendering typically requires a large number of images. The goal of this research is to minimize the number of training images required by identifying optimal camera positions for image capture, as shown in Figure 3. Previously, obtaining images from desired positions was challenging due to the difficulty in obtaining accurate camera positions. In this study, we addressed this issue by integrating the training process with the Unity environment. In this study, a camera within the virtual environment was used to capture object images, creating custom datasets along with the camera's transformation information for NeRF training. The NeRF Image Selector significantly enhanced training efficiency by introducing a novel pseudo-label. This pseudo-label, functioning like a reinforcement learning reward, enabled a self-supervised learning approach that improved NeRF model performance in nearly all cases.

Adaptive Image Synthesis with Denoising Diffusion Probabilistic Models This research focused on training separate diffusion models for object and background generation, and then synthesizing their outputs to create cohesive images that integrate both elements, as shown in Figure 2. developed an innovative 4-channel learning approach for diffusion models, where the object diffusion model learns not only the RGB channels but also an additional mask channel during training. This enables the model to generate a mask output, which is then used to merge the object and background.

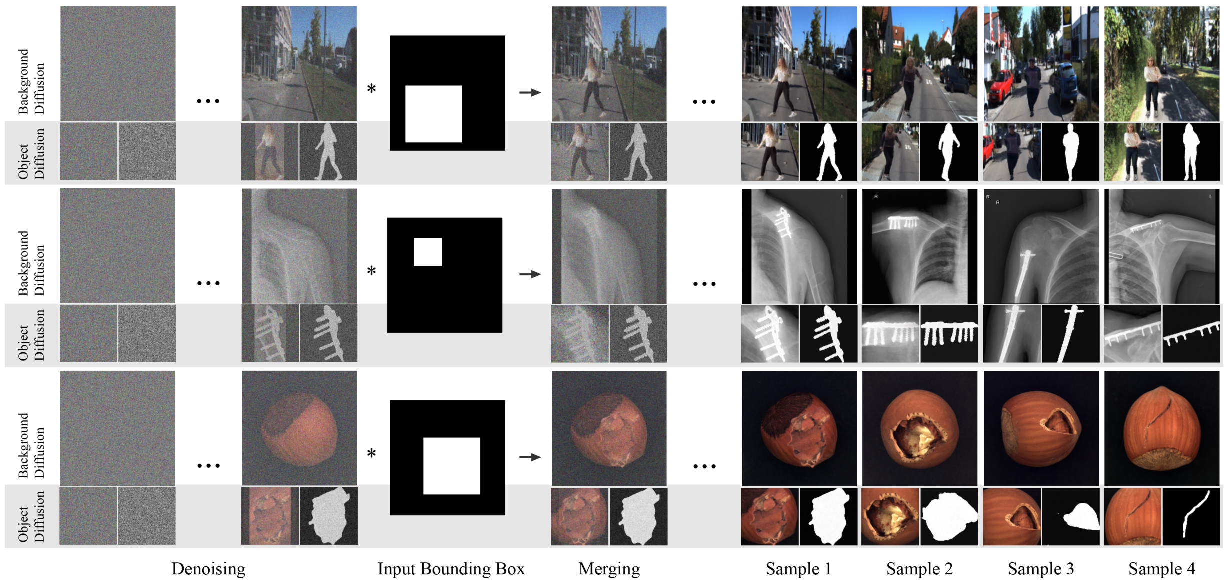

Figure 2: Two Denoising Diffusion Probabilistic Models (DDPM) are used, one of which is trained with 4-channel (RGB+Mask). The first row illustrates the fusion of outputs from two diffusion models: the background model trained on the KITTI dataset and the object model trained on the TikTok dataset. The second row shows a case where the background diffusion model was trained on shoulder X-ray images without implants, and the object diffusion model was trained on implant objects. The third row uses the hazelnut subset of the MVTec dataset, where the background diffusion model was trained on normal hazelnut images, and the object diffusion model was trained on anomalies.

To achieve more natural blending between the outputs of the two models, I utilized iterative re-sampling and de-noising during the diffusion process, which significantly enhanced the overall composition. Additionally, objects were placed within specified bounding box locations, providing precise control over their positioning. This approach enabled the model to autonomously generate and optimize a mask for synthesis, eliminating the need for manually defined masks. The results demonstrated strong performance across highly semantically diverse datasets, achieving high-quality image synthesis while reducing reliance on human input for mask creation.

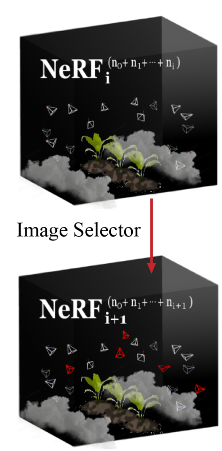

Figure 3: Visualization of iterative NeRF model training. Here, n_i represents the number of images selected before the i-th training iteration. Initially, n_0 is 0 or randomly chosen.

Future Research Agenda

An outstanding AI system necessitates the harmony of diverse AI domains. Currently, AI research spans a vast range of specialized areas, each advancing the boundaries of what's achievable across various domains. As AI tackles increasingly complex real-world challenges, the field grows in diversity and depth. **My future research focuses on effectively integrating these diverse components to manage the growing complexity of AI systems.** To address this, I aim to develop methods that leverage cross-disciplinary knowledge to enable quick understanding and adaptability. This foundation provides the vision to integrate the right elements into intricate collaborations, creating systems that are robust and scalable. By using insights drawn from diverse experiences, I aim to fill the "black box" of AI solutions with transparent, innovative approaches, illuminating paths that were previously hidden. My ultimate goal is to design AI systems that **harmonize diverse domains**, achieving perfect synergy and unlocking new possibilities for real-world applications.