УДК 004.896

SYNTHETIC DATASET GENERATION FOR UNDERWATER OBJECT DETECTION USING DIFFUSION MODELS

Bassel Alshawareb (ITMO University)
Scientific supervisor – Dr. Of Science, Professor, Sergey Kolyubin (ITMO University)

Introduction. Object detection is an important task for multiple underwater robotic applications, including ecological monitoring and infrastructure inspection. Despite the advancements of object detection in terrestrial environments, known methods and models exhibit a performance drop when used underwater due to inherent challenges such as light absorption and scattering. Therefore, developing domain transfer approaches for DNN models is in demand with training and validation datasets being major enablers. This work proposes a novel approach to address the challenge of task-oriented underwater datasets scarcity by generating prompt-guided realistic synthetic images, and subsequently fine-tuning common object detectors on this data.

Main part. Diffusion models have demonstrated a remarkable capability for generating realistic synthetic data [1]. By utilizing a relatively small amount of data, it is possible to fine-tune the model to generate images from the same distribution adapted for the domain of interest, which is underwater in our case.
This work proposes a two-stage process: an image generator training stage, and an inference stage, when automated domain-specific image generation is happening. In the first stage, real underwater images, as well as their captions, are used to fine-tune latent diffusion models using LoRA [2], while image captions are generated using BLIP2 model [3]. After training the generator model, text prompts are prepared for generating underwater scenes with objects of interest. Object categories used in these prompts are selected from the ImageNet dataset [4].
To automate checking prompt-to-image matching and image regeneration process, this work proposes a verification block. To do so this block uses BLIP2 to obtain a caption of the generated image and then measure cosine similarity between embeddings created for this caption and the input prompt. If the metric is below a threshold, the image is declined and the generation process with the same prompt is started again with a different seed. The final result of this process is synthetic datasets adapted for the underwater domain having similar ground truth information that ImageNet dataset provides, which makes it suitable for object detection tasks.
This dataset can be used later to adapt existing detectors to underwater environments by fine-tuning the backbones to learn the representations of the challenging light conditions, and therefore enhance the mean average precision of detection.

Conclusion. This work introduces an effective approach for generating synthetic data tailored to the underwater domain using diffusion models for enhancing underwater object detection and fine-tuning the backbone of the detector on it. It proposes a pipeline for automatically generating underwater images with an automated verification step. Future work will be on exploring

different methods for utilizing the generated underwater dataset for enhancing object detection performance.

References

1. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-resolution image synthesis with latent diffusion models // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. — 2022. — P. 10684–10695.
2. Akbarian Bafghi R., Bagwell C., Ravichandran A., Shrivastava A., Raissi M. Fine Tuning without Catastrophic Forgetting via Selective Low Rank Adaptation // arXiv e-prints. — 2025. — arXiv–2501.
3. Li J., Li D., Savarese S., Hoi S. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models // International conference on machine learning. — PMLR. 2023. — P. 19730–19742.
4. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. - CVPR. 2009. - P.. 248-255.