Image enhancement based on multi-scale attention network

Dudu Liu, Yiping Huang, Haojun Tan, Guanlin zhou

I. INTRODUCTION

MUTI-SCALE Attention Network (MAN) is based on Convolutional Neural Network (CNN) and is used for single image super-resolution (SR) tasks. With the development of technology, people's demand for image clarity is increasing. However, due to various reasons, the clarity of the images we obtain always fails to meet our requirements. Therefore, we hope to use a multi-scale attention network, which was presented at the Conference on Computer Vision and Pattern Recognition this year, to enhance the image and improve its resolution.

II. METHODOLOGY

The core methodology of our project revolves around the implementation of a Multi-Scale Attention (MAN) for enhancing the resolution and quality of images and videos. This approach leverages the strengths of deep learning, particular Convolutional Nerual Network (CNN), to address the challenges associated with low-resolution and poor lighting conditions commonly found in image and video data.

Data acquisition and preprocessing:

- Data Collection: Collect vehicle and pedestrian datasets online.
- Dataset construction: We will build a dataset containing images and videos of different resolutions and qualities. This dataset will include images and videos captured under different lighting conditions, with a particular focus on low light scenes.
- Preprocessing: The dataset will undergo preprocessing steps to normalize the data. This includes resizing images to a uniform size, decomposing video frames, and increasing data diversity and robustness through data augmentation techniques such as rotation, scaling, and flipping.

Model building:

- Using MAN: We will use the multi-scale attention network (MAN) proposed in the paper. This network is capable of capturing long-range dependencies and local details through its Multi Scale Large Kernel Attention (MLKA) and Gated Spatial Attention Units (GSAU).
- Customized Enhancement: The MAN model will be customized and adjusted to meet specific requirements for image and video enhancement. Including adjusting the network to focus on tasks such as noise reduction, resolution enhancement, and dynamic range expansion.

Model training:

- Define loss function: We will use a combined loss function that includes mean square error (MSE) and perceptual loss to ensure that the model not only minimizes pixel level errors, but also maintains the perceptual quality of enhanced images and videos.
- Optimization: When training the model, try using various optimizers and adopt a learning rate plan that decays over time to effectively fine tune the model weights.

Data post-processing:

- Quality assurance: The enhanced images and videos will undergo a series of post-processing steps for refinement. This includes denoising, color correction, and contrast enhancement to ensure that the final output meets high-quality standards.

Written on October 30, 2024.