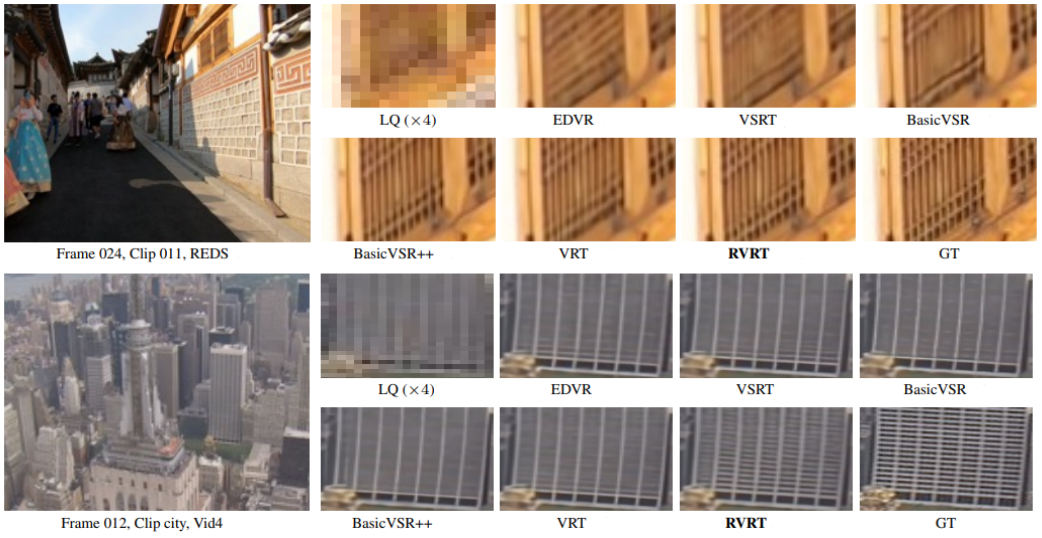
Trimble / Bilberry : AI Engineer technical exercise
Part 2 : Paper Review

Chappe Aslan
August 2023

Recurrent Video Restoration Transformer
with Guided Deformable Attention

The Recurrent Video Restoration Transformer (RVRT) introduces a novel approach to video restoration that blends elements of parallel and recurrent methods to achieve state-of-the-art performance in various restoration tasks, including video super-resolution, deblurring, and denoising. Leveraging the transformer architecture, widely successful in natural language processing, RVRT tackles complex challenges in computer vision.

Video restoration poses a significant challenge due to the temporal and spatial intricacies of videos. Existing methods can be categorized as parallel or recurrent based on their parallelizability. Parallel methods estimate all frames simultaneously, while recurrent methods propagate features from one frame to another sequentially, accumulating information from previous frames for the restoration of subsequent ones.



RVRT consists of three main components: shallow feature extraction, recurrent feature refinement, and high-quality frame reconstruction. For feature extraction, a convolutional layer is employed to extract features from the low-quality video. Recurrent feature refinement is achieved through feature refinement modules that leverage the temporal correlation between frames. Additionally, RVRT introduces a Guided Deformable Attention (GDA)

module to align neighboring video clips and facilitate feature aggregation across time. The GDA module utilizes pre-alignments based on optical flow to adaptively adjust features.

Experimental results demonstrate RVRT's superiority over existing methods in various video restoration tasks, including super-resolution, deblurring, and denoising. Compared to competing models, RVRT not only offers better performance but also features a smaller model size, lower testing memory consumption, and shorter runtime. Nevertheless, what particularly captures my interest in this publication are the promising avenues it opens for leveraging the capabilities of RVRT to enhance the precision of data processed by the model. The prospect of harnessing this technology to improve outcomes across various domains seems highly appealing.

Source

Liang, J., Fan, Y., Xiang, X., Ranjan, R., Ilg, E., Green, S., Cao, J., Zhang, K., Timofte, R., & Gool, L.V. (2022). Recurrent Video Restoration Transformer with Guided Deformable Attention. *ArXiv, abs/2206.02146*.

https://github.com/JingyunLiang/RVRT
https://arxiv.org/pdf/2206.02146.pdf