Final Project: Image Enhancement Effect on Text-to-Video Retrieval

Group Members:
Zahra Dehghanian
Parsa Haghighi
Elham Abolhassani
M. Taha Teimuri

Image Processing
CE409331
Professor: Dr. Kasaei

August 2023

1 Introduction

In this project, we investigate the effects of various image enhancement techniques on a text-to-video retrieval task. Our focus is to understand how enhanced images might contribute to the effectiveness and precision of the retrieval process. By using the MSR-VTT dataset and implementing the CLIP model in a zero-shot setting, we hope to generate new insights about the correlation between image quality and retrieval accuracy.

2 Problem Definition

The main objective of this project is to study the influence of different image enhancement algorithms on a text-to-video retrieval task. The nature of the problem involves improving the image quality before the retrieval process and subsequently comparing the results against the non-enhanced scenario. The challenge lies in measuring the variation in retrieval accuracy caused by these enhancements and interpreting the implications of these variations.

2.1 Image Enhancement Techniques

In this section, we aim to improve the quality of digital images by employing various image enhancement techniques. Specifically, we utilize techniques such as Histogram Equalization, Gaussian Blurring, Contrast Limited Adaptive Histogram Equalization, sharpening, and adjusting images in the HSV mode. These techniques are widely used to remove noise, increase contrast, optimize brightness levels, and enhance overall image quality.

Histogram Equalization is a well-known technique used in computer image processing to improve contrast in images. It redistributes the pixel intensities of an image, effectively spreading out the most frequent intensity values and stretching out the intensity range of the image. This method increases the global contrast of images and allows areas of lower local contrast to gain higher contrast.

Gaussian Blurring, on the other hand, is a technique used to smooth images by removing high-frequency noise. It involves convolving the image with a Gaussian kernel, which reduces the sharp transitions and details in the image, resulting in a smoother appearance.

Contrast Limited Adaptive Histogram Equalization (CLAHE) is a variant of histogram equalization that enhances local contrast in images. Unlike ordinary histogram equalization, CLAHE computes several histograms, each corresponding to a distinct section of the image, and uses them to redistribute the lightness values of the image. This technique is particularly useful for enhancing details in specific regions of an image while preserving overall contrast.

Sharpening is a technique used to enhance edges and details in an image. It involves increasing the contrast along edges, making them appear more pronounced and enhancing the overall clarity of the image.

Finally, adjusting the image in the HSV (Hue, Saturation, Value) mode involves converting the image from the BGR (Blue, Green, Red) color space to the HSV color space, adjusting the hue, saturation, and value components, and then converting the image back to the BGR color space. This allows for more precise control over color and brightness adjustments in the image.

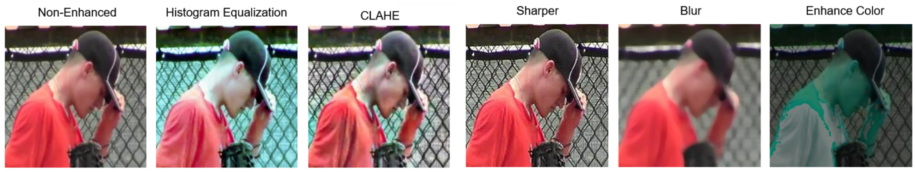

Figure 1: Sample of Enhancment algorithms

2.2 Improving CLIP feature vectors by extracting text parts

In this section we use different ocr techniques to extract text parts in each frame in order to improve quality of CLIP generated feature vectors. In our experiments, we saw that the generated CLIP feature vectors are very sensitive to presence of text in frames i.e., if we add some text to a frame, the CLIP feature vector of this frame changes a lot while the content of these two frames are the same. So, inorder to slove this issue, we extract text parts of frames with two ocr techniques and after that by bluring, pixelating or replacing these parts with their corresponding parts in the previous frame, we try to imporve generated feature vectors.

Easyocr and Kerasocr are two ocr techniques that we used them in this part. Easyocr is a python library that allows you to easily extract text from images and it is built on top of the well-known ocr engines Tesseract and Kraken. Kerasocr is also a powerful open-source library that provides a high-level api for ocr tasks.

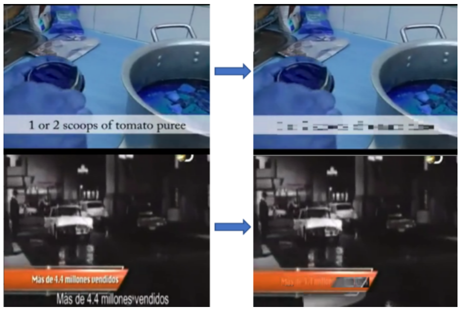
1 or 2 scoops of tomato puree
Mas de 4,4 millones vendidos
Mas de 4,4 millones vendidos

Figure 2: Sample of Text Enhancement

2.3 Text-to-Video Retrieval

Text-to-video retrieval involves using a text-based query to search and retrieve relevant video content. It's an essential component of this project as it enables us to measure the efficacy of the image enhancement techniques. Using a zero-shot setting with the CLIP model, we analyze how these enhancements affect the retrieval process, specifically regarding the accuracy and relevance of the retrieved video content

3 Pipeline

Our approach utilizes a zero-shot setting to observe the effect of different image enhancement techniques on frames to enhance retrieval metrics. In our pipeline, we first enhance the image frames using the mentioned techniques. We then uniformly sample 'k' frames from each video. These frames are processed through the CLIP model to generate feature vectors which are compared against the feature vectors of the text queries. The retrieval process is then conducted based on these comparisons.

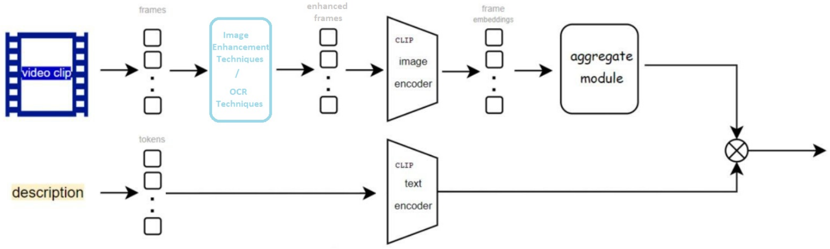

Figure 3: Our Pipeline

4 Results

In this section, we discuss the results obtained from our expriments. We have used 5 Image Enhancing techinques and they have results varying from much better to much worse then the None Enhanced imaged. One of the first things to notice is that algorithms that increase the details in image tend to have better result like sharpening but at the same time if the details are enhanced too much we may get worse results. Also Obviously techiniques that removes details from image like bluring will worsen the results.

Method	**R@1↑**	**R@5↑**	**R@10↑**	**Median R↓**	**Mean R↓**
Non-Enhanced	12	36	60	17.2	20.1
OCR	10.0	**41.9**	**69.9**	17.3	20.2
Histogram Equalization	12	24	46	18.15	20.08
CLAHE	8	26	46	17.8	20.0
Sharper	**12.1**	40	68	**15.6**	**18.2**
Too sharp	10.0	28.0	64.0	16.6	19.2
Blur	2.0	12	25	23.1	23.0
Enhance Color	8	22	33.9	18.2	21.0

Table 1: Text to Video recall values

5 Conclusion

By doing this expriment we can concolude that Having good image enhancing algorithms can drastticly increase the accuracy of a Text to Video model.