Text to video generation for News Stories

Problem Statement

Text to Video generation is the problem of converting a given text to video automatically.Previous work on the generative relationship between text and video has focused on producing text captioning from video. However, the inverse problemof producing videos from text has is a challenging problem for existing methods.In this research the problem is tackled by training a conditional generative model to extract both static and dynamic information from text. This is done by having a Variational Autoencoder (VAE)and a Generative Adversarial Network (GAN). The static features, called "gist," are used to sketch text-conditioned back-ground color and object layout structure. Dynamic features are considered by transforming input text into an image filter. To obtain a large amount of data for training the deep-learning model, we develop a method to automatically create a matched text-video corpus from publicly available online videos.The challenges which need to be addressed are the selection of key terms which need to be highlighted in video,extracting the gist from text and showing them in video,only showing the highlights and that too in meaningful sentence form.

Background Work

Despite of the challenges given in the problem statement plant disease detection is still an active area of research. Numerous approaches have been proposed over the years.Recent work on video generation has decomposed video into a static background, a maskand moving objects.One approach to combining the text and gist information is to simply concatenate the feature vectors from the encoded text and the gist, as was previously used in image generation.Video generation is intimately related to video prediction.Video prediction focuses on making object motion realisticin a stable background. Recurrent Neural Networks (RNNs)and the widely used sequence-to-sequence model. Acommon thread among these works is that a convolutionalneural network (CNN) encodes/decodes each frame andconnects to a sequence-to-sequence model to predict thepixels of future frames. In addition, (Liu et al.2017) proposed deep flow networks for video-frame interpolation.There is also significant work on video generation conditioned on a given image.In these works, it is important to distinguishpotential moving objects from the given image. In contrastto video prediction, these methods are useful for generating a variety of potential futures, based upon the currentimage.

Another approach for detection and differentiation of plant diseases can be achieved using Support Vector Machine algorithms. This technique was implemented for sugar beet diseases, depending on the type and stage of disease, the classification accuracy was between 65% and 90%.Another approach based on leaf images and using ANNs as a technique for an automatic detection and classification of plant diseases was used with *K*-means as a clustering procedure. ANN consisted of 10 hidden layers.

Methodology

1. **Data Collection and Dataset Preparation**: Large number of videos downloaded from YouTube for each keyword along with their title,description and tag.Clean videos from the Kinetics Human Action Video Dataset are also used
2. **Training:** Training the deep convolutional neural network for making video of news out of text description model is done. For making video out of text news the use of keywords extracted from text news is done.
3. **Testing:**In this phase it used entirely different test set of news article for prediction.

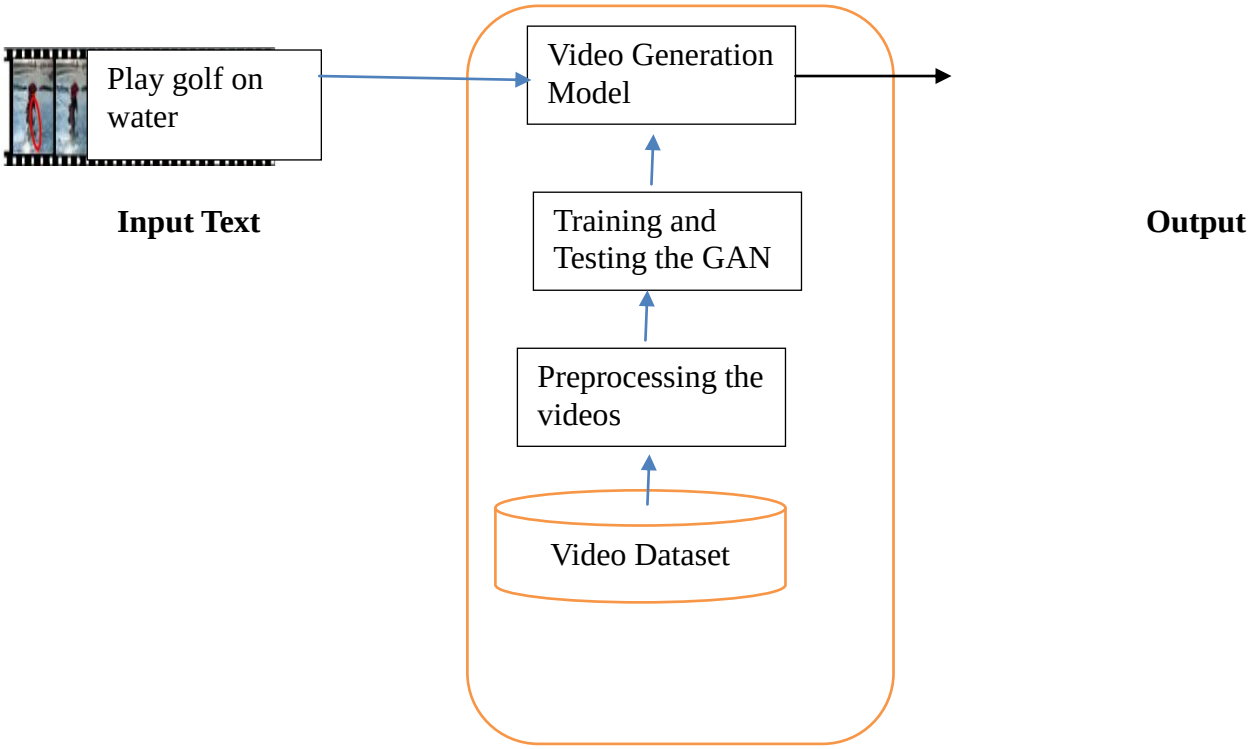

Figure 1: Architecture of Text to Video Generation System

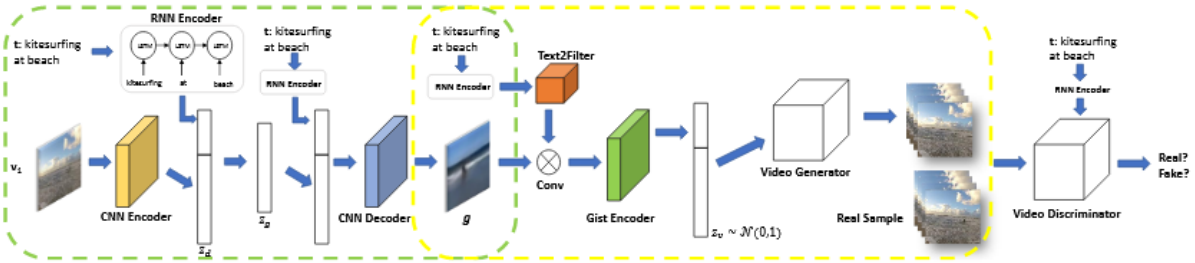

Figure 2: Framework of Text to Video Generation System showing the components of the system[Yitongi Li et al. Video Generation from Text 1st October 2017]

Experimental Design

Dataset: Large number of videos downloaded from YouTube for each keyword along with their title,description and tag.Clean videos from the Kinetics Human Action Video Dataset are also used

Evaluation Measures:Measures such as accuracy and precision of the content played needs to be considered

Software and Hardware Requirements:Python based Deep Learning libraries like Keras Training will be conducted on NVIDIA GPUs