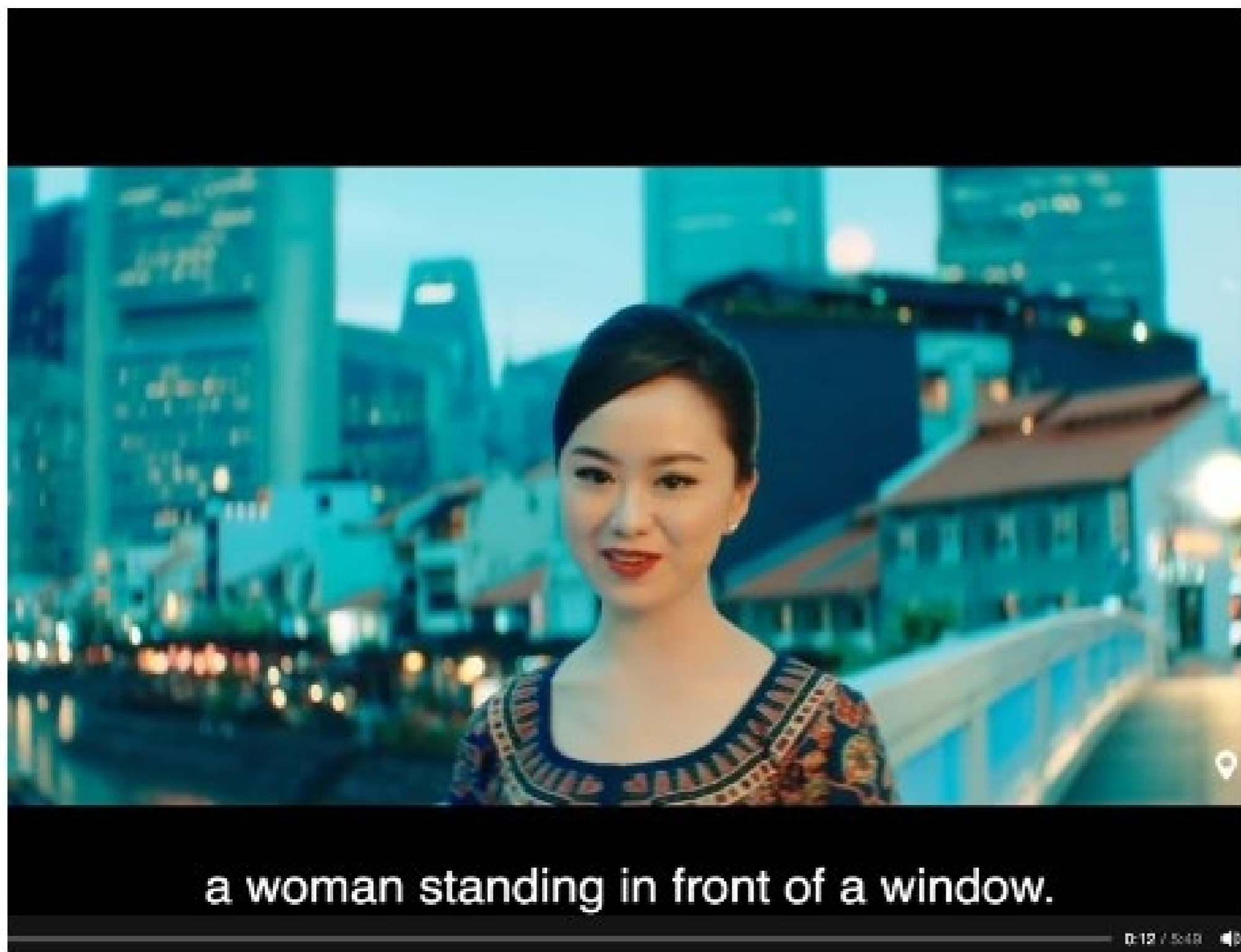




Image Captioning: Textual Description For Images From Videos

Student: Neo Jia Jun

Supervisor: Assoc Professor Chng Eng Siong



Caption [im2txt]: a group of people standing in front of a bus .
Caption [Camera2Caption]: a bus is parked on the side of the street .
Caption [Image Captioning with Visual Attention]: a group of people walking down a street.
Caption [NeuralTalk2]: a group of people standing in front of a bus

Technologies Used:

- Deep Learning (CNN, LSTM)
- FFMPEG
- Bash and Python Scripts
- VGGNet, ResNet

Using deep learning technologies such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM). The video is feed into a in-house processing system using FFMPEG to split the videos into images. The images are then processed in the models and captions are then generated. A Graphic User Interface (GUI) is created to allow users to select the best caption to be used in the video.