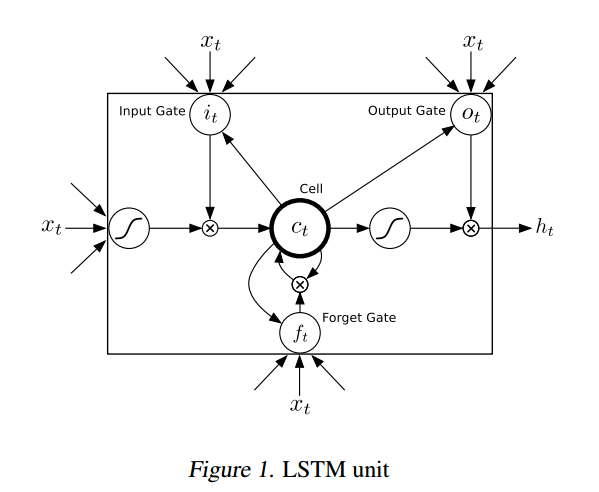
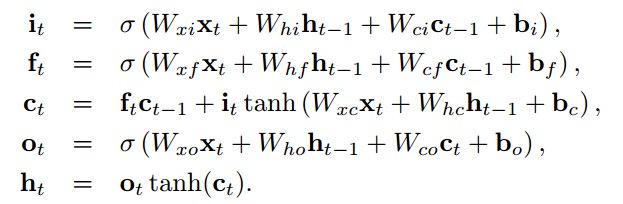
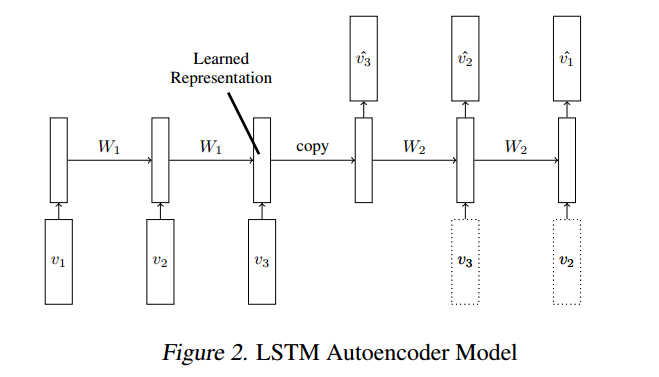
Unsupervised Learning of Video Representations using LSTMs

1. Tổng quan
   1. Sử dụng multilayer LSTM network để học biểu diễn video
   2. Model:
      * Encoder LSTM ánh xạ chuỗi đầu vào thành mộ biểu diễn có kích thước cố định.
      * Decoder LSTM (single or multiple) thực hiện recontrucsting đầu vào hoặc là dự đoán chuỗi tương lai.
   3. Thực nghiệm trên 2 loại input:
      * Tập các ảnh pixels
      * High-level representation của video được trích xuất từ mạng convolutional đã được train từ trước.
2. Approach
   * + Sử dụng LSTM encoder-decoder để học video representation
     + Encoder LSTM: nhận một chuỗi khung hình và đưa ra một biểu diễn.
     + Biểu diễn này được decode qua LSTM khác để tạo chuỗi mục tiêu
     + Có 2 loại decoder:
       - Biểu diễn lại đầu vào
       - Dự đoán chuỗi tương lai
3. Model Description
   1. Long short term memory

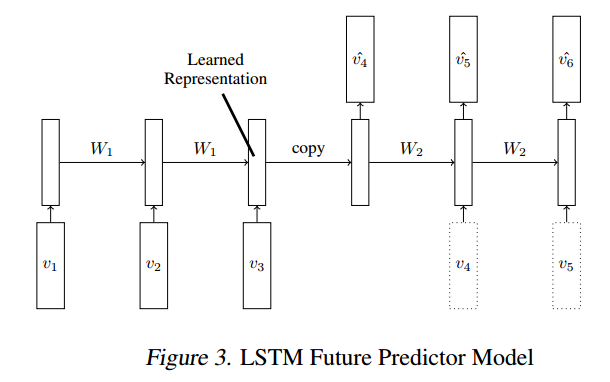




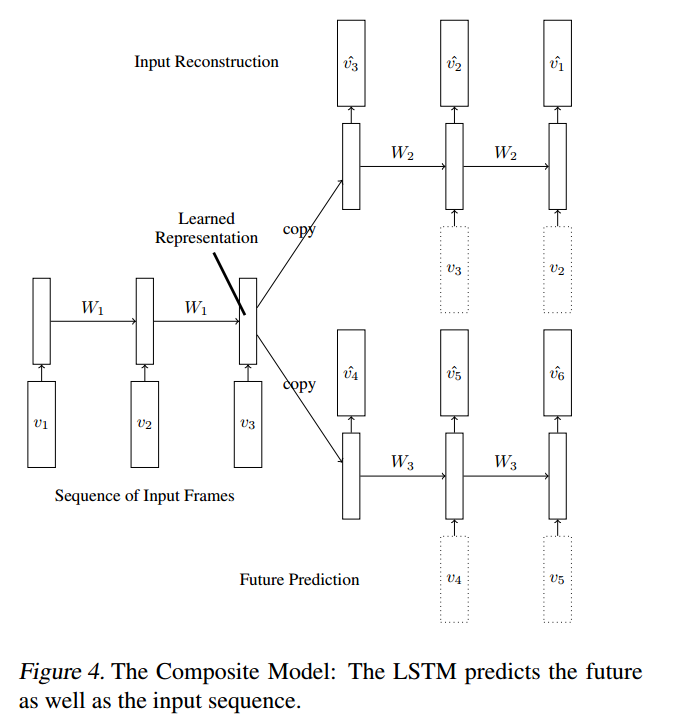
* 1. LSTM Autoencoder Model



* Mô hình bao gồm 2 mạng RNN (LSTM): encoder LSTM và decoder LSTM
* Input của mô hình dạng vector chuỗi các giá trị liên tiếp
* Encoder LSTM nhận chuỗi đầu vào và deocder LSTM nhận kết quả từ encoder và đưa ra dự đoán cho chuỗi mục tiêu.
* Chuỗi mục tiêu là chuỗi input nhưng có thứ tự đảo ngược.
* Decoder có 2 loại :
  + Có điều kiện: decoder nhận các giá trị output gần nhất trước đó làm đầu vào.
  + Không có điều kiện: deocoder không nhận đầu vào.
* Kiến trúc mô hình có thể mở rộng ra nhiều layers LSTM
  1. LSTM Future Predictor Model



* Cấu trúc tương tự Autoencoder Model, ngoại trừ đầu ra của decoder không phải chuỗi input mà là chuỗi dự đoán tiếp theo của input.
  1. A Composite Model



* Mô hình là kết hợp của cả 2 task: reconstructing the input và predicting the future
* Encoder được yêu cầu phải đi đến trạng thái phù hợp cho cả predict future và reconstruct input
* Mô hình khắc phục được các nhược điểm của các mô hình trên:
  + Autoencoder: cần phải ghi nhớ tất cả các thông tin input, nó không có ích cho dự đoán tương lai.
  + LSTM Future Predictor Model: chỉ cần ghi nhớ các thông tin cuối của input ( nó có ý nghĩa hơn cho việc dự đoán tương lai). Nhưng nếu muốn dự đoán tất cả chuỗi input thì không thể chỉ chú ý đến các giá trị cuối.

1. Experiments
   * + Phần này họ chỉ nêu các kết quả thực nghiệm trên các mô hình nên em không trình bày trong đây.
     + Cơ bản thì composite model đạt kết quả tốt nhất trong các mô hình đã nêu.