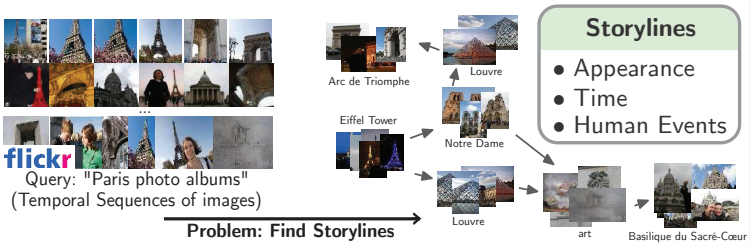
Learning Visual Storylines with Skipping Recurrent Neural Networks

Gunnar A. Sigurdsson, Xinlei Chen, Abhinav Gupta

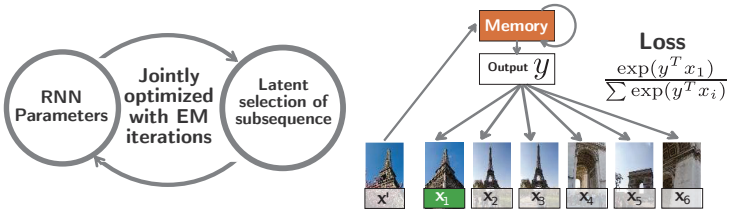
Carnegie Mellon University — The Robotics Institute

Language Technologies Institute

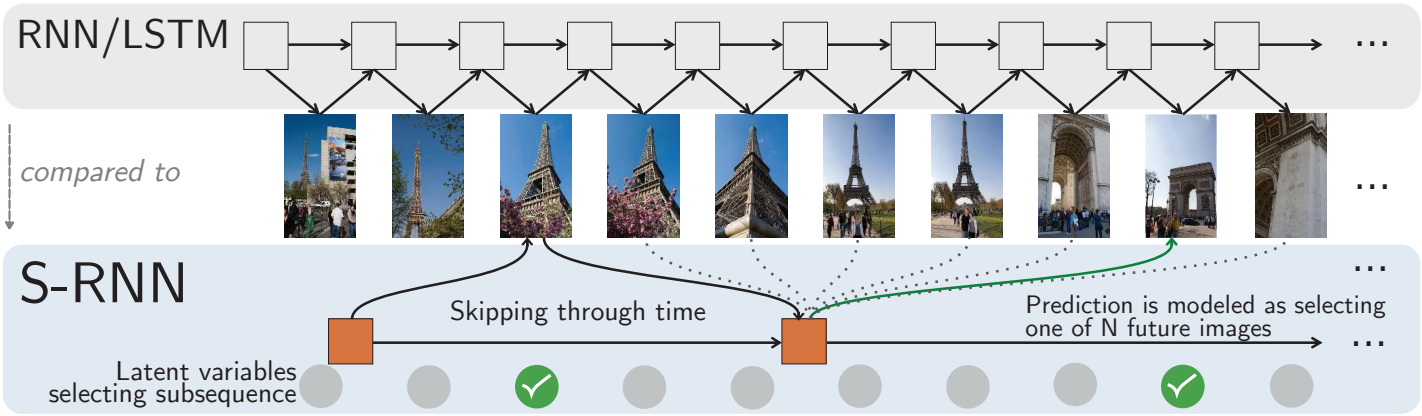


- We learn the ***visual and temporal*** parts of concepts: Wedding, Christmas
- Learn in an ***unsupervised*** way with recurrent networks
- Our S-RNN learns the ***common long-term latent stories***
- We train one model for each ***concept*** (e.g. Paris)

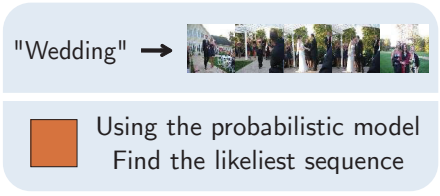
Method

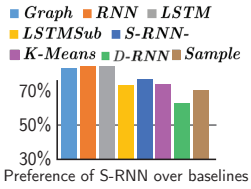


- ***Train RNN over all possible subsequences***
- Rewritten as sequential update equations for E and M steps
- Fits in the RNN pipeline: ***Infer next variable, then backprop***
- Prediction is ***softmax over future images***
  - Images are represented as their fc7 features



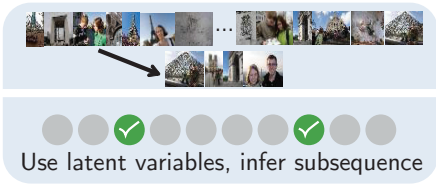
Storylines

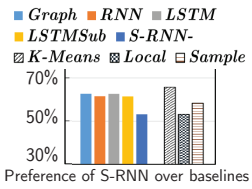




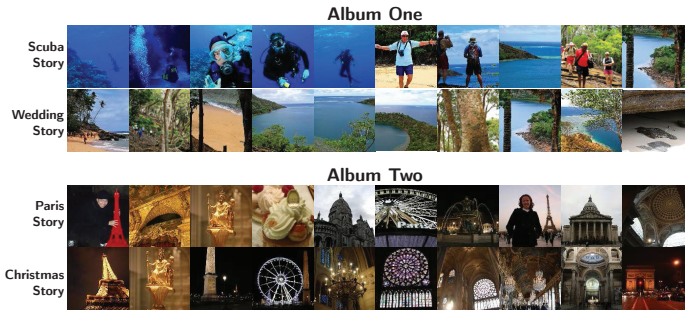
- Finding a image sequence that captures a concept

Summarization

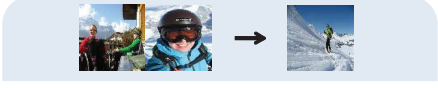




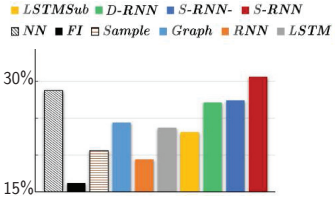
- Find a short summary of a photo album
- Inferring latent variables gives a summary
- Album can be summarized in various ways:

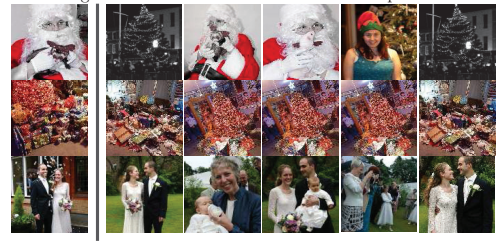


Prediction



- Predicting what comes next in a photo album





github.com/gsig