

Pre-Meeting - Thursday, June 20th at 1PM

Peyman Bateni

June 20, 2019

1 Admin (ETC: 2-3 minutes):

- Currently also working on "Sentimentalized Visual Captioning" for submission to WACV (deadline is Friday July 26th at midnight GMT). I'm mainly brushing stuff up and designing the cross-dataset evaluation metric for the second section that focuses on adding adjectives to captions, but once Professor Sigal is back, I'll finish up the write up (my goal is to be done with this by first week of July).
- I'm also talking to people regarding the directed studies (have talked to Adam: negative response; currently talking to Christian.) to supervise my directed studies in a structural format that would minimize the work. My goal is to start on it early in July, so I'm able to finish things up early.
- Updates regarding my efforts with Inverted AI are as noted by the Slack chat and the details will be delivered in person. I'm also helping out with some stuff on the self-driving project albeit haven't been able to add much too recently.
- I'm also dealing with a very painful infected hair root on the back of my neck which has unfortunately resulted in two rather unproductive days this week as I was receiving treatments and recovering. [t]

2 Technical Updates (ETC: 10-12 minutes)

- The problem definition is (very much) complete and can be viewed on the Github repository. There are clarifications to be made along with an opening to look more into active learning. We'll have to go through this once more.
- I've done additional reading on existing active learning and semi-supervised learning models and architectural designs. There's much more reading to do with respect to extending the work to detection as it seems to be the case that unlabelled examples lose their usefulness in detection while still adding value in the classification case.
- A top-down generative model of the problem is shown in Figure 1. The notational details of the model are as follows:
 - $C_i \in C_{allclasses}$ described the class for the i th sampled object
 - $P_i \in Z^5$ described the position of the object, with the first 3 values describing the actual position of the object in a 3 dimensional world with (0, 0, 0) defining the bottom left corner at depth 0. The other two values describe the rotation to the top and to the right of the object respectively. Zero rotation occurs when the object plane is parallel to the frame.
 - The parameters for $VRNN(\theta_1)$, $LSTM(\theta_2)$ and the GAN are learned from data and few-shot detection is to be performed using inference.

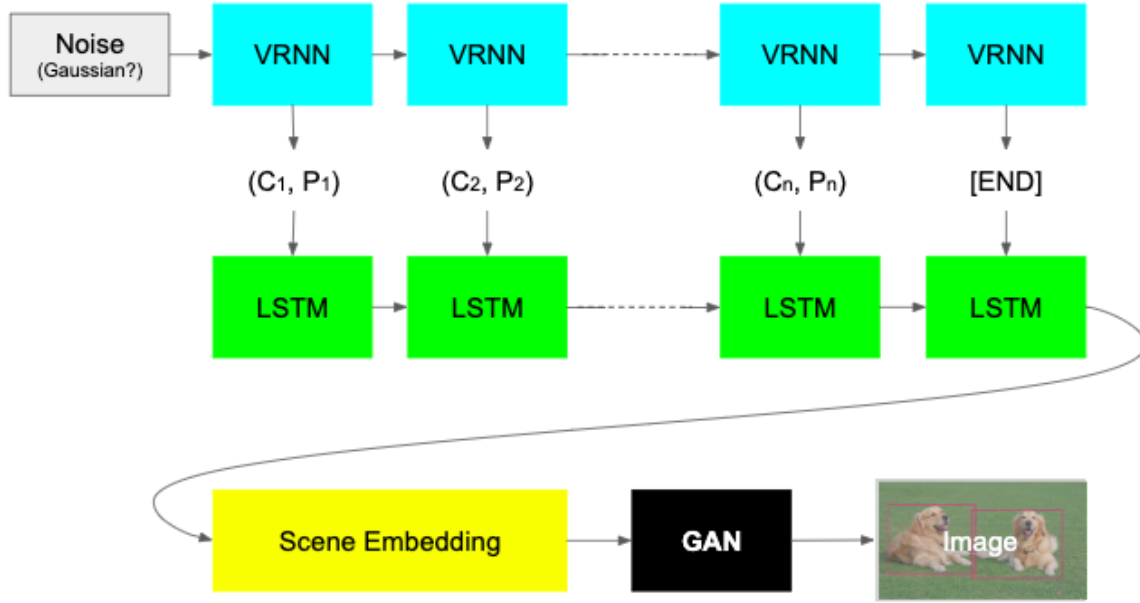


Figure 1: Top down generative model for few-shot visual detection.

The issues involving the model are as follows:

- $VRNN(\theta_1)$ must be invariant to object order (unless otherwise we are able to define a sensible ordering)
- Inference needs to be fast (potential use of deep neural networks instead for detection?)
- Current model assumes availability of labelled data only. How can we incorporate unlabelled data?
- Should the model have the option to request labels for a select few, how should we select what examples to get labels for? Based on estimated loss improvement? Based on decision boundaries in a multi-dimensional knowledge space?
- Lastly, how can we potentially incorporate multi-modal information in form of transfers from class definition graphs or an NLP corpus?
- As for more achitectoral solutions, here are five ideas we’ve been looking at so far:
- Five ideas I suggested last time:
 - Visual GloVe
 - Contextual graph matching
 - Unbounded hierarchical nonparametric bayes (was rejected due to the exclusion of label discovery as part of the problem definition)
 - Transfer functions with mixture of gaussian boundary definitions
 - Generalizable AIR (motivation for the current top-down model)

3 Goals for Today (ETC: hopefully 45 minutes, likely much more)

- Review problem definition
- Discuss top-down model's design and issues involved including inference, order invariance, and inclusion of unlabelled data.
- Discuss other architectural solutions that we've looked out.
- Plan future reading strategies, works I should evaluate, etc.