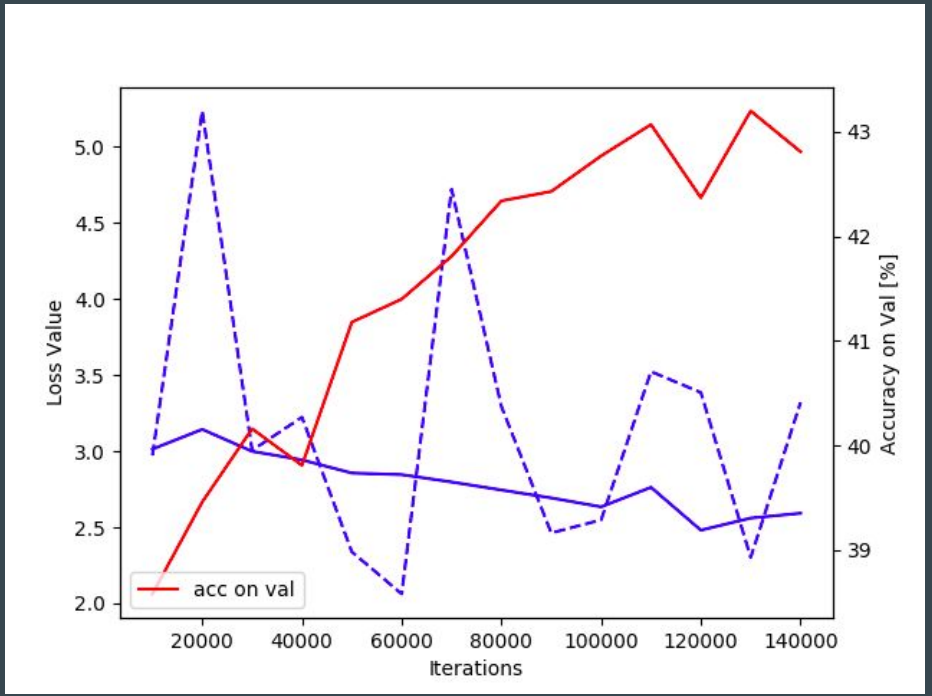
Visual Question Answering



Gera Versfeld

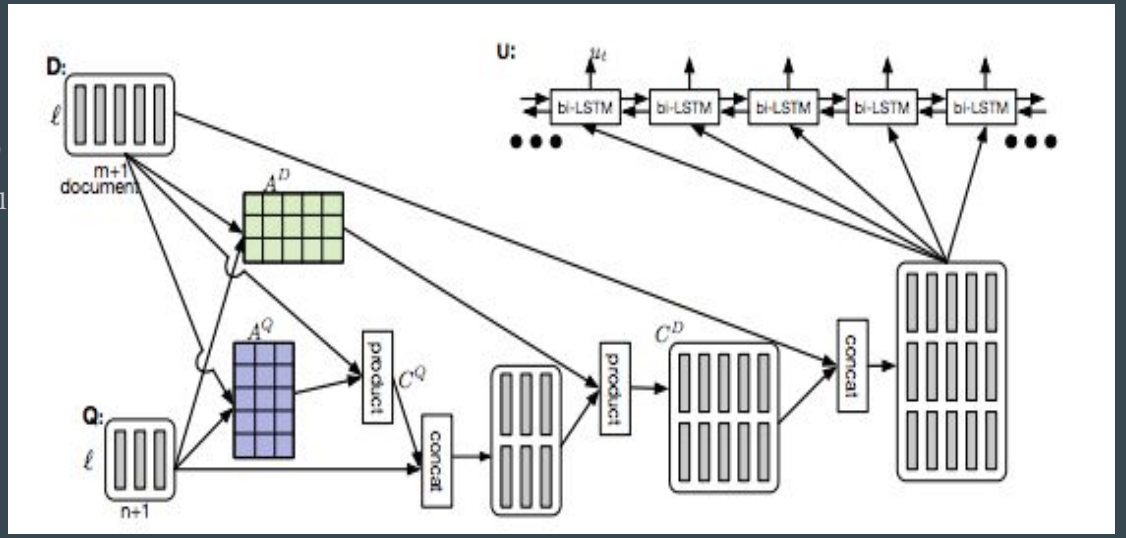
VQA -challenge

- Submitted MCB baseline with POS weighting - 60.81% accuracy on test split
- Training MCB baseline on different features to make ensemble (inception-v3 image features and word2vec question embeddings)



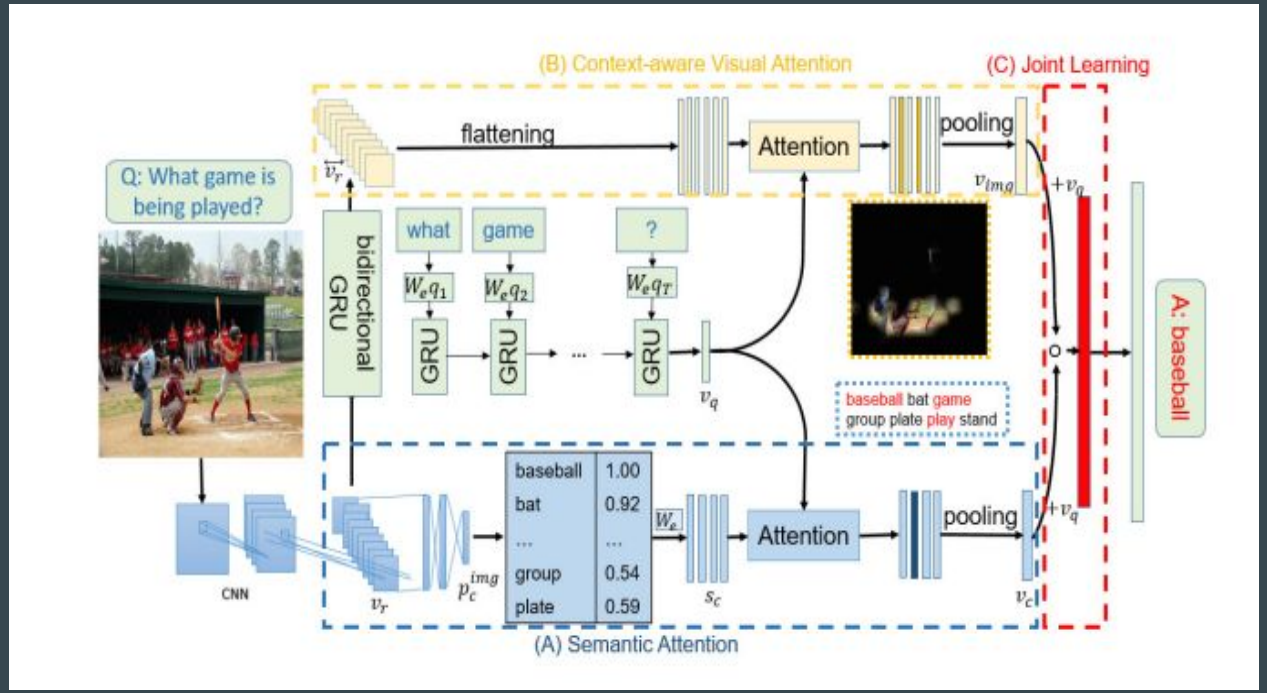
Dynamic Co-Attention Networks for Question Answering[1]

- Co-attention encoder:
  - Affinity matrix:
    - $A^Q = \text{softmax}(L) \in R^{(m+1(n+1)}$
    - $A^D = \text{softmax}(L^T) \in R^{(n+1(m+1}$
  - summary/attention contexts
    - $C^Q = D^*A^Q \in R^{l^*(n+1)}$
    - $C^D = [D; Q^Q]\ A^D \in R^{l^*(n+1)}$

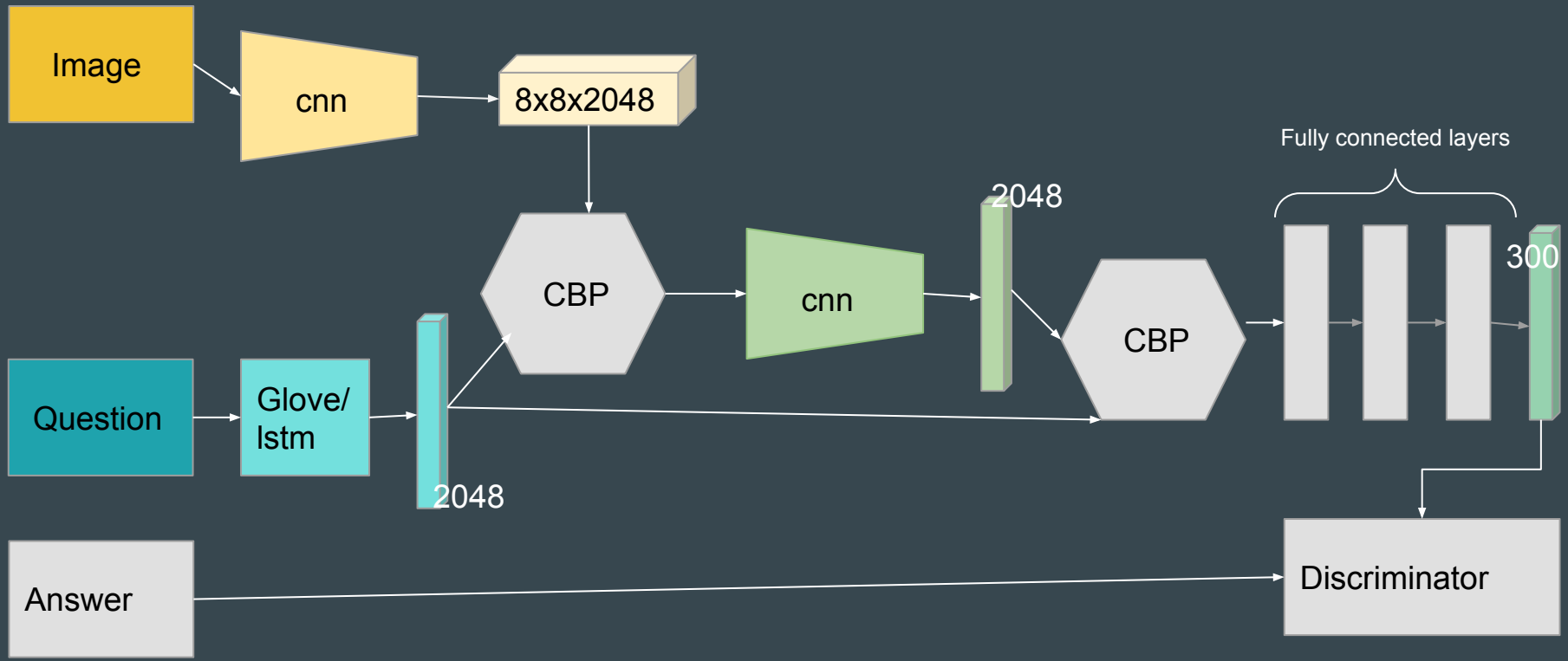


Multi-level Attention Networks for Visual Question Answering[2]

- Extract abstract/ high level image features
- Semantic attention
- Less memory/ computation



My proposed model: regression/GAN



Where we are at now:

- Training with different features for ensemble
- Implemented generator part of GAN model

Goals for the week:

- Implement discriminator
- Continue training my model
- Evaluate ensemble on test-dev