

11-777 Report 2: Baselines and Model Proposal

First Last 1* First Last 2* First Last 3* First Last 4*
{ID1, ID2, ID3, ID4}@andrew.cmu.edu

1 Baseline Models and Metrics (2 pages)

Please explain the baselines you are using. This includes (but is not limited to):

1. Unimodal baselines
2. Very simple multimodal models
3. Alternate choices for modules (e.g. encoders/decoders)

(Explain all your choices and what types of interactions exist. What can a simple detector figure out of about the task? What can an LM infer from the prompt? What if I just pass detections to a model... – you implement 2*N baselines – these can use pretrained encoders/detectors)

1.1 Unimodal Baselines

We include the following unimodal baselines: ... we expect these to capture ... (these are implemented by the group)

1.2 Simple Multimodal Baselines

We include a very simple multimodal baseline that ... ignores history... doesn't have attention ... doesn't require pretraining (these are implemented by the group)

1.3 Competitive Baselines

We run N competitive baselines available at public repos. These include: (these are models you found in the literature, github, etc with existing checkpoints – no training. You should be able to run these so you can do analysis on them in R3. Additionally, they need to be rerun if you've done your own data sampling)

1. System 1 is a ... whose key insight is ...

2. System 2 is a ... whose key insight is ...

3. System 3 is a ... whose key insight is ...

*Everyone Contributed Equally – Alphabetical order

Methods	Dev	
	Accuracy \uparrow	L_2 Error \downarrow
Unimodal 1 ()		
Unimodal 2 ()		
Unimodal 3 ()		
Simple Multimodal 1 ()		
Simple Multimodal 2 ()		
Simple Multimodal 3 ()		
Previous Approach 1 ()		
Previous Approach 2 ()		
Previous Approach 3 ()		

2 Results (1 page)

Replace columns with the correct metrics for your task (extrinsic). What are all the things you can measure, and why are they beneficial?

Metric 1

Metric 2

Metric 3

3 Model Proposal (1 page)

Include a diagram (e.g. labeled flow chart) of all modules. This is not final!

3.1 Overall model structure

3.2 Encoders

Describe encoders for each modality and at least one alternatives for each. Explain the relative strengths of each option (e.g. coverage, efficiency, ...)

3.3 Decoders

3.4 Loss Functions

Describe both your primary task loss and three possible auxiliary losses that might improve performance. Justify your choices.

4 Team member contributions

Member 1 contributed ...

Member 2 contributed ...

... contributed ...