

Baselines and Analysis

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

| Methods | Dev | |
|------------------------|--------------|-------------------|
| | Detections ↑ | Invalid Actions ↓ |
| Unimodal 1 () | | |
| Unimodal 2 () | | |
| Unimodal 3 () | | |
| Simple Multimodal 1 () | | |
| Simple Multimodal 2 () | | |
| Simple Multimodal 3 () | | |
| Previous Approach 1 () | | |
| Previous Approach 2 () | | |
| Previous Approach 3 () | | |

1 Analysis (4 pages)

This section should include plots. For example, how key metrics vary with a specific hyperparameter, task complexity, etc.

1.1 Intrinsic Metrics

These are not the task itself (they might overlap with auxiliary losses) but are skills the model should have.

Intrinsic Metric 1

Intrinsic Metric 2

Intrinsic Metric 3

*Everyone Contributed Equally – Alphabetical order

1.2 Qualitative Analysis and Examples (full page table)

A full page table that includes: Example input, comparison of model outputs, theory about reason for failure. Every model presented should have at least one failure condition

Insights Two paragraphs on what you found and how they impact your expectations for the main model, or how they have inspired you to change your proposed approach, etc

2 Team member contributions

Member 1 contributed ...

Member 2 contributed ...

... contributed ...