SANDWICH Supplementary MATERIAL

March 14, 2025

1 Neural Network Setting

Table 1: Neural Network Setting

Component	**Parameter**	**Value**
MARL training	Neural network layers	2
	Neurons per layer	128
	Learning rate	0.001
	Initial noise	0.75
	Noise decay	0.999995
	Minimum noise	0.01
	Terminate reward (**I**)	10

2 User Study Interface

3 Weighted Fleiss' Kappa Calculation

Definition of the Weight Matrix

Since ratings are ordinal, we apply a quadratic weight function to penalize large discrepancies more than small ones. The weight matrix w_{ij} is defined as:

$$w_{ij} = 1 - \frac{(i-j)^2}{(k-1)^2} \tag{1}$$

where:

- i, j are rating categories $(1, 2, \ldots, k)$.
- k is the total number of rating categories.
- The further apart the categories, the smaller the weight.

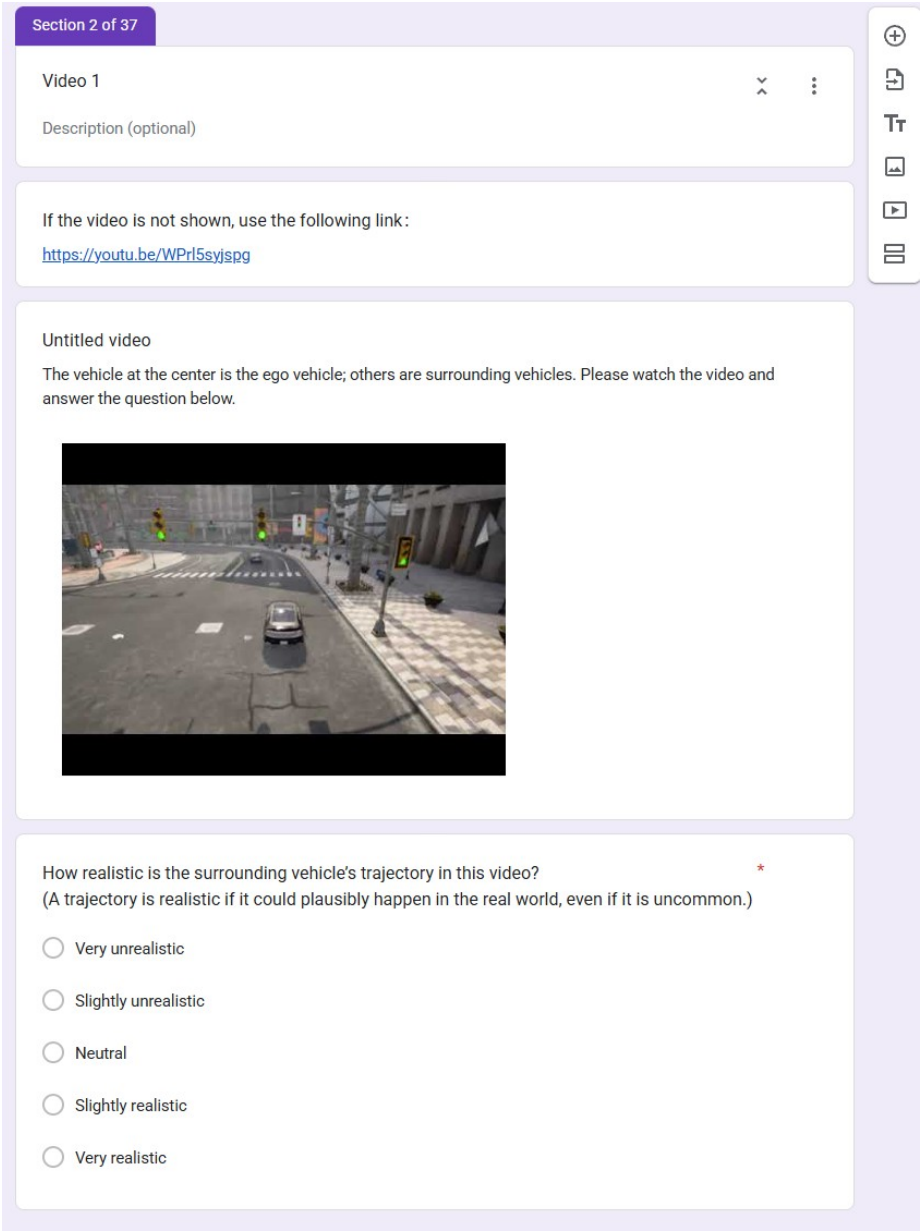

Figure 1: User Study Interface

Observed Agreement

For each subject (e.g., a video being rated), let f_{ij} be the number of raters assigning a rating of j. The total number of ratings for a given subject i is:

$$p_i = \sum_{j=1}^{k} f_{ij} \tag{2}$$

The observed agreement for subject i is calculated as:

$$P_i = \frac{\sum_{j=1}^{k} \sum_{l=1}^{k} w_{jl} \cdot f_{ij} \cdot f_{il}}{p_i(p_i - 1)} \tag{3}$$

where:

- f_{ij} is the count of raters giving category j to subject i.
- $p_i(p_i - 1)$ is the total number of rating pairs.
- If $p_i = 1$, then P_i is set to 0 (as no agreement is possible).

The overall observed agreement across all subjects is:

$$\bar{P} = \frac{1}{N} \sum_{i=1}^{N} P_i \tag{4}$$

where N is the total number of subjects.

Expected Agreement

The proportion of ratings assigned to each category across all subjects is:

$$p_j = \frac{\sum_{i=1}^{N} f_{ij}}{\sum_{i=1}^{N} p_i} \tag{5}$$

The expected agreement, assuming ratings were randomly assigned, is:

$$P_e = \sum_{j=1}^{k} \sum_{l=1}^{k} w_{jl} \cdot p_j \cdot p_l \tag{6}$$

Computation of Weighted Fleiss' Kappa

The final formula for the weighted Fleiss' Kappa is:

$$\kappa_w = \frac{\bar{P} - P_e}{1 - P_e} \tag{7}$$

where:

- $\bar{P}$ is the observed agreement.
- P_e is the expected agreement.
- If $\bar{P} = P_e$, then $\kappa_w = 0$, indicating no agreement beyond chance.

Interpretation of Weighted Fleiss' Kappa

The interpretation of κ_w is as follows:

- $0.00 - 0.20$: Slight agreement
- $0.21 - 0.40$: Fair agreement
- $0.41 - 0.60$: Moderate agreement
- $0.61 - 0.80$: Substantial agreement
- $0.81 - 1.00$: Almost perfect agreement

For example, if we compute:

$$\kappa_w \approx 0.714 \tag{8}$$

this suggests **substantial agreement** among raters.