1. More Comparisons with DLow/LDS

In Tab. 1 we give more comprehensive comparisons of our stimulus verification against sampling approaches, *i.e.* DLow and LDS, using both GAN-based model (Social GAN) and CVAE-based model (GroupNet) as base prediction models. From the table we can see that our proposed stimulus verification consistently outperforms both DLow and LDS, proving its effectiveness. Further, it is worth noting that as a post-prediction process, stimulus verification is not in conflict with the sampling methods mentioned above, meaning that it is possible to combine them for even better performances.

Method	ETH/UCY	GCS	NBA
Social GAN	0.61 / 1.21	6.06 / 8.98	1.78 / 2.48
+*DLow*	0.54 / 1.05	5.69 / 7.64	1.71 / 2.32
+*LDS*	0.52 / 1.00	5.65 / 7.56	1.69 / 2.27
+*Ours*	0.49 / 0.94	5.29 / 6.95	1.59 / 1.94
GroupNet	0.25 / 0.44	3.76 / 5.20	1.13 / 1.26
+*DLow*	0.24 / 0.42	3.66 / 5.04	1.11 / 1.21
+*LDS*	0.23 / 0.39	3.57 / 4.85	1.09 / 1.18
+*Ours*	0.23 / 0.37	3.50 / 4.58	1.08 / 1.12

Table 1. Comparison between stimulus verification and other sampling approaches.

2. Contribution and Analysis of NMS

In Tab. 2, we show the results of three base models on the ETH/UCY dataset without using NMS, from which we can see declination on the performances, proving its necessity. Further, to demonstrate the NMS's effectiveness on increasing the diversity of final outputs, we visualize in Fig. 1 examples on final outputs of these base models before and after adopting NMS. From the figure we can see that the final outputs tend to be homogeneous without NMS. Yet after applying NMS, the diversities of such outputs increase significantly.

Method	Social GAN	PCCSNet	GroupNet
w/o NMS	0.49 / 0.97	0.23 / 0.46	0.27 / 0.48
w/ NMS	0.49 / 0.94	0.20 / 0.39	0.23 / 0.37

Table 2. Comparison between stimulus verification with & without NMS on ETH/UCY Benckmark.

3. Inference Time

The average inference time of our social verifier, context verifier, DLow and LDS is about 16, 120, 70, 4 ms/task respectively. As for reference, the speed of GroupNet (only
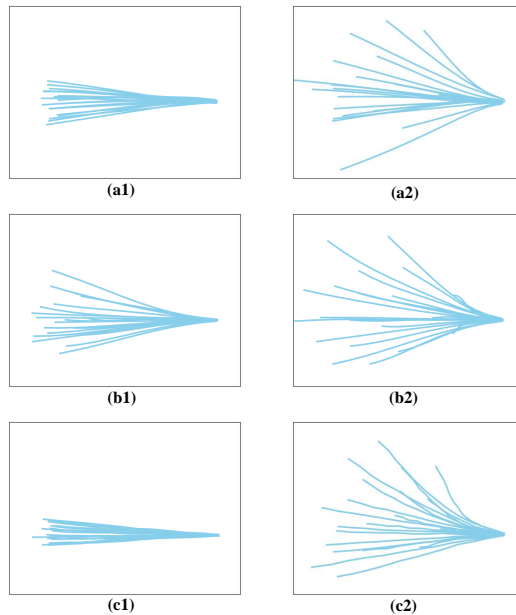

Figure 1. Illustration of final outputs of Social GAN(a), PCC-SNet(b) and GroupNet(c) before (left) and after (right) applying NMS.

social as stimulus) is about 40 ms/task and the speed of Sophie (both social and context as stimuli) is about 160 ms/task. The speed of the verifier mainly depends on the stimulus feature extractor.

4. Limitations

We discuss the limitations of stimulus verfication as follows. First, the verifier is currently designed as a probabilistic model optimized via MLE, which would require a large amount of data for training, otherwise the learned verifier may be heavily biased. Second, in order to perform the stimulus verification, extra candidate samples need to be predicted first and brings computational overhead. Luckily, this can be done in parallel to avoid increasing much inference time.