

Supplementary Material for the Paper

Conformal Prediction for Semantically-Aware Autonomous Perception in Urban Environments

A	Proofs	12
A.1	Background For Proofs: Cascaded Conformal Prediction	12
A.2	Proof of Theorem 3.1	12
A.3	Proof of Corollary 3.1	13
B	Structure of the Knowledge Graph	14
C	Results for Further Task Sequences	17
D	Qualitative Results	17
D.1	Qualitative Results on the ROAD Dataset	18
D.2	Qualitative Results on the Waymo/ROAD++ Dataset	20

A Proofs

A.1 Background For Proofs: Cascaded Conformal Prediction

Cascaded conformal prediction (cascaded-CP) [25] is a technique that allows to prune prediction sets sequentially for a *single task*, using a cascade of different non-conformity scores over m steps. Since different statistical tests are applied to the data set, the multiple hypothesis testing (MHT) problem arises, which leads to an increased family-wise error rate (i.e., false positives), making the CP procedure invalid. Cascaded-CP makes use of p-value correction procedures M , such as Simes corrections, to account for MHT problem. Cascaded-CP is formalized in Theorem A.1.

Theorem A.1 (Cascaded-CP [25]) *For any sequence of non-conformity measures (S_1, \dots, S_m) , which yields p-values (P_1, \dots, P_m) and $\alpha \in [0, 1]$, the prediction set $C^j(X_{test})$ at step $j < m$ is defined as:*

$$C^j(X_{test}) = \{Y \in \mathcal{Y} : \tilde{P}_j^y > \alpha\} \quad (6)$$

where \tilde{P}_j^y is the corrected p-values using the procedure M for candidate y at step j . Then $\forall j \in [1, m]$, $C^j(X_{test})$ satisfies Equation 1, and $C^m(X_{test}) \subseteq C^j(X_{test})$.

In this work, we show that Theorem A.1 can be extended to cover multiple tasks in cascade. Furthermore, in our case the tasks are not required to share the same label space. By constructing a knowledge graph facilitating semantic mapping between tasks, our approach accommodates different tasks. Notably, Cascaded-CP can be seen as a sub-case of our work if all tasks are identical.

A.2 Proof of Theorem 3.1

We consider 2 tasks \mathcal{T}_c and \mathcal{T}_l that are performed sequentially. Our goal is to prove that the prediction set C_l^{KRPS} obtained by performing any CP procedure on the set C_l^K , which represents the

semantic mapping $\mathcal{M}_{c \rightarrow l}$ of the set C_c^{KRPS} , satisfies 2 properties: marginal coverage, and semantic consistency with respect to C_c^{KRPS} and \mathcal{K} .

Marginal Coverage First, we prove the marginal coverage property of the set C_l^{KRPS} , that is:

$$\mathbb{P}[Y_{test}^l \in C_l^{KRPS}(X_{test})] \geq 1 - \alpha \quad (7)$$

The set C_c^{KRPS} is constructed using a CP procedure, meaning that it satisfies Equation 1, and we have:

$$\mathbb{P}[Y_{test}^c \in C_c^{KRPS}(X_{test})] \geq 1 - \alpha \quad (8)$$

The semantic mapping $\mathcal{M}_{c \rightarrow l}$ is a deterministic mapping that assigns a set of possible locations to each element of $C_c^{KRPS}(X_{test})$. Knowing that the true value of the subsequent task Y_{test}^l is the image of the true value of the starting task Y_{test}^c , the resulting mapping set $C_l^{\mathcal{K}}(X_{test}) = \mathcal{M}_{c \rightarrow l}(C_c^{KRPS}(X_{test}))$ contains the Y_{test}^l with a probability that is at least equal to $1 - \alpha$, as expressed in Equation 9.

$$\mathbb{P}[Y_{test}^l \in C_l^{\mathcal{K}}(X_{test})] \geq 1 - \alpha \quad (9)$$

From the CP coverage Theorem 2.1, we can conclude the existence of a p-value $P_l^{\mathcal{K}}$ that satisfies:

$$\mathbb{P}[Y_{test}^l \in C_l^{\mathcal{K}}(X_{test})] \geq 1 - \alpha \iff \mathbb{P}[P_l^{\mathcal{K}} \leq \alpha] \leq \alpha \quad (10)$$

At this step, we have two distinct p-values for the subsequent task \mathcal{T}_l , namely $P_l^{\mathcal{K}}$ and P_l . The p-value $P_l^{\mathcal{K}}$ is employed in the construction of the set $C_l^{\mathcal{K}}$, whereas P_l is utilized for forming the set $C_l^l(X_{test})$, representing the CP-based set for task \mathcal{T}_l independently of the knowledge provided by the prior task \mathcal{T}_c , established on C_c^{KRPS} .

Subsequently, we execute a cascaded-CP procedure using Theorem A.1 on the p-values $P_l, P_l^{\mathcal{K}}$. We incorporate a p-value correction procedure denoted as M , resulting in corrected p-values $\hat{P}_l, \hat{P}_l^{\mathcal{K}}$, and we have:

$$\begin{aligned} \mathbb{P}[Y_{test}^l \in C_l^l(X_{test}) \cap C_l^{\mathcal{K}}(X_{test})] &\geq 1 - \alpha \\ \iff \mathbb{P}[Y_{test}^l \in C_l^{KRPS}(X_{test})] &\geq 1 - \alpha \end{aligned}$$

which is the result stated in Equation 3.

Semantic Consistency Our goal now is to show that the newly constructed prediction set C_l^{KRPS} is semantically consistent with respect to C_c^{KRPS} and \mathcal{K} . This result of semantic consistency comes from the fact that for each element Y^l of C_l^{KRPS} , we have $Y^l \in C_l^l(X_{test}) \cap \mathcal{M}_{c \rightarrow l}(C_c^c(X_{test}))$, meaning that $C_l^{KRPS} \subseteq \mathcal{M}_{c \rightarrow l}(C_c^c(X_{test}))$, which gives the semantic consistency property.

Implications of Theorem 3.1 The direct implication of Theorem 3.1 is that we can further refine the prediction sets generated by any CP procedure given \mathcal{K} and a related task, by removing classes that are not semantically consistent with other tasks, without losing the property of marginal coverage, provided that the p-values are properly corrected. The semantic refinement in KRPS plays an additional role in highlighting corner cases. In urban applications, it is typical to start with a basic knowledge graph and incrementally add new class relationships as data becomes available. This process, however, may encounter corner cases or semantic inconsistencies, like vehicles on sidewalks, not covered by the knowledge graph. KRPS addresses these instances by outputting an empty prediction set when a semantically consistent vehicle position cannot be found. This empty set signals potential knowledge graph gaps or corner cases, prompting further investigation by end-users. Updating the knowledge graph and performing a calibration step is sufficient to adapt to new data without retraining the model.

A.3 Proof of Corollary 3.1

The sets C_c^{KRPS} , C_l^{KRPS} and, C_a^{KRPS} are 3 prediction sets constructed using KRPS, in the described order. Given this, our goal is to prove that C_a^{KRPS} is semantically consistent with respect

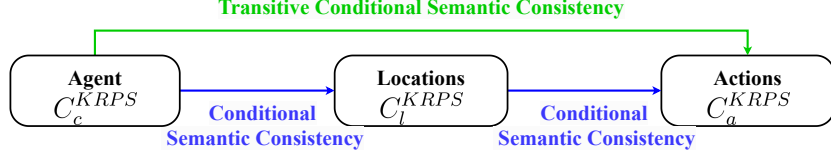


Figure 3: Implications of Theorem 3.1 (blue) and Corollary 3.1 (green) on the semantic consistency.

520 to C_c^{KRPS} and \mathcal{K} . Since C_a^{KRPS} is constructed based on Theorem 3.1, we have C_a^{KRPS} is seman-
 521 tically consistent with respect to C_l^{KRPS} and \mathcal{K} , i.e.,

$$\forall Y_a \in C_a^{KRPS}, \exists Y_l \in C_l^{KRPS} / Y_a \in \mathcal{M}_{l \rightarrow a}(Y_l) \quad (11)$$

522 Since C_l^{KRPS} is constructed using KRPS based on Theorem 3.1, C_l^{KRPS} is semantically consistent
 523 with respect to C_c^{KRPS} and \mathcal{K} and we have:

$$\forall Y_l \in C_l^{KRPS}, \exists Y_c \in C_c^{KRPS} / Y_l \in \mathcal{M}_{c \rightarrow l}(Y_c) \quad (12)$$

524 Based on Equations 11 and 12, we can establish that:

$$\forall Y_a \in C_a^{KRPS}, \exists Y_c \in C_c^{KRPS} / Y_a \in \mathcal{M}_{c \rightarrow a}(Y_c) \quad (13)$$

525 Which gives that C_a^{KRPS} is semantically consistent with respect to C_c^{KRPS} and \mathcal{K} .

526 **Implications of Corollary 3.1** Corollary 3.1 establishes the transitive properties of semantic con-
 527 sistency, as defined using Definition 3.1. As depicted in Figure 3, this result implies that two pre-
 528 diction sets from sequential tasks maintain semantic consistency, even if constructed independently,
 529 as long as they are built in sequence relative to a shared task in the middle. This finding holds sig-
 530 nificance as knowledge bases evolve over time with the inclusion of new tasks. Employing KRPS
 531 ensures guarantees of semantic consistency with all prior tasks, by uniquely verifying semantic con-
 532 sistency for the latest executed task.

533 B Structure of the Knowledge Graph

534 In the following, we provide more details about the knowledge graph used to model the semantic
 535 relationships between the entities in the urban environment. We adopt a simple, yet effective on-
 536 tological model based on the following semantic relationship: *Agent* performs *action* in *location*.
 537 It is possible to adopt different task orders, e.g., *Action* is performed in *location* by *agent*. In our
 538 setup, the classification of an agent, its location, and its action, correspond to different tasks, that are
 539 performed by separate models or separate heads of a single model. This is to ensure that we consider
 540 a multitask setup, in contrast to situations where a model outputs a triplet, which we consider as a
 541 case of multi-class classification tasks.

542 We adopt the class labels provided by the ROAD and the Waymo/ROAD++ datasets³ [1, 24] for
 543 all the tasks that we consider. For Completeness, we report the list of agent, location, and action
 544 classes, as they are described in the ROAD dataset, in Table 2, Table 3, and Table 4, respectively.

545 Based on the class labels for each task, the semantic mapping functions $\mathcal{M}_{T_i \rightarrow T_j}$, where T_i and T_j
 546 represent the start and the subsequent tasks, respectively, are constructed through the examination
 547 of possible label assignments between the tasks in the training set. Table 5 summarizes the semantic
 548 mappings $\mathcal{M}_{c \rightarrow a}$ and $\mathcal{M}_{c \rightarrow l}$. The semantic mappings $\mathcal{M}_{a \rightarrow c}$ and $\mathcal{M}_{a \rightarrow l}$ are represented in Table 6.
 549 Finally, the semantic mappings $\mathcal{M}_{l \rightarrow c}$ and $\mathcal{M}_{l \rightarrow a}$ are represented in Table 7.

³Waymo/ROAD++: <https://sites.google.com/view/road-plus-plus/dataset>

Label Name	Abbreviation
Car	Car
Medium Vehicle	MedVeh
Large Vehicle	LarVeh
Bus	Bus
Motorbike	Mobike
Emergency Vehicle	EmVeh
Pedestrian	Ped
Cyclist	Cyc
Vehicle Traffic Light	TL
Other Traffic Light	OthTL

Table 2: List of agent classes and their abbreviations as reported in the ROAD dataset [1].

Label	Abbreviation
In vehicle lane	VehLane
In outgoing lane	OutgoLane
In incoming lane	IncomLane
In outgoing cycle lane	OutgoCycLane
In incoming cycle lane	IncomCycLane
On left pavement	LftPav
On right pavement	RhtPav
On pavement	Pav
At junction	Jun
At crossing	Xing
At bus stop	BusStop
At parking	parking

Table 3: List of used location classes and their abbreviations as reported in the ROAD dataset [1].

Label	Abbreviation
Moving away	MovAway
Moving towards	MovTow
Moving	Mov
Braking	Brake
Stopped	Stop
Indicating left	IncatLft
Indicating right	IncatRht
Hazard lights on	HazLit
Turning left	TurnLft
Turning right	TurnRht
Moving right	MovRht
Moving left	MovLft
Overtaking	Ovtak
Waiting to cross	Wait2X
Crossing road from left	XingFmLft
Crossing road from right	XingFmRht
Crossing	Xing
Pushing object	PushObj
Traffic light red	Red
Traffic light amber	Amber
Traffic light green	Green

Table 4: List of used action classes and their abbreviations as reported in the ROAD dataset [1]

Agent	List of Actions	List of Locations
Ped	MovAway,MovTow,Mov,Stop,Wait2X,XingFmLft,XingFmRht,	VehLane,IncomLane,Pav,LftPav,RhtPav,Jun,xing,BusStop
Car	MovAway,MovTow,Brake,Stop,IncatLft,IncatRht,TurLft,TurRht	VehLane,OutgoLane,IncomLane,Jun
Cyc	MovAway,MovTow,Stop,TurLft,XingFmLft	VehLane,OutgoLane,OutgoCycLane,IncomLane,IncomCycLane
Mobike	MovAway,MovTow,Brake,Stop,IncatLft,IncatRht,TurLft,TurRht	VehLane,OutgoLane,IncomLane,Jun
MedVeh	MovTow,Stop,TurRht, TurLft, Brake	IncomLane,Jun,OutgoLane
LarVeh	MovTow,Stop,TurRht, TurLft, Brake	IncomLane,Jun,OutgoLane
Bus	MovTow,Stop,XingFmLft	VehLane,IncomLane,Jun, BusStop

Table 5: The semantic mappings between the agent classes and the action classes ($\mathcal{M}_{c \rightarrow a}$), and the agent classes and the location classes ($\mathcal{M}_{c \rightarrow l}$).

Action	List of Agents	List of Locations
MovAway	Ped,Car,Cyc,MedVeh,Bus,LarVeh	VehLane,OutgoLane,OutgoCycLane,Pav,LftPav,RhtPav,Jun
MovTow	Ped,Car,Cyc,MedVeh,Bus,LarVeh	VehLane,IncomLane,IncomCycLane,LftPav,RhtPav,Jun
Mov	Ped	Pav
Brake	Car	VehLane,Jun
Stop	Ped,Car,Cyc,MedVeh,Bus	VehLane,IncomLane,IncomCycLane,Pav,LftPav,RhtPav,Jun,BusStop
IncatLft	Car	VehLane,Jun
IncatRht	Car	IncomLane,Jun
TurLft	Car,Cyc	VehLane,Jun
TurRht	Car,MedVeh	IncomLane,Jun
Ovtak	Car	VehLane
Wait2X	Ped	LftPav,RhtPav
XingFmLft	Ped,Car,Cyc,Bus	VehLane,IncomLane,Jun,xing
XingFmRht	Ped	VehLane,IncomLane,RhtPav,Jun
Xing	Ped,Cyc	Xing
PushObj	Ped	LftPav,RhtPav

Table 6: The semantic mappings between the action classes and the agent classes ($\mathcal{M}_{a \rightarrow c}$), and the agent classes and location classes ($\mathcal{M}_{a \rightarrow l}$).

Location	List of Agents	List of Actions
VehLane	Ped,Car,Cyc,Bus	MovAway,MovTow,Brake,Stop,IncatLft,TurLft,XingFmLft,XingFmRht
OutgoLane	Car,Cyc	MovAway
OutgoCycLane	Cyc	MovAway
IncomLane	Ped,Car,Cyc,MedVeh,Bus	MovTow,Stop,IncatRht,TurRht,XingFmLft,XingFmRht
IncomCycLane	Cyc	MovTow,Stop
Pav	Ped	MovAway,Mov,Stop
LftPav	Ped,Cyc	MovAway,MovTow,Stop,Wait2X,PushObj
RhtPav	Ped	MovAway,MovTow,Stop,Wait2X,XingFmRht,PushObj
Jun	Ped,Car,Cyc,MedVeh,Bus	MovAway,MovTow,Brake,Stop,IncatLft,IncatRht, TurLft, TurRht, XingFmLft, XingFmRht
xing	Ped	XingFmLft
BusStop	Ped,Bus	Stop
parking	Car	parking

Table 7: The semantic mappings between the location classes and the agent classes ($\mathcal{M}_{l \rightarrow c}$), and the location classes and action classes ($\mathcal{M}_{l \rightarrow a}$).

Task	Score	Method	$\alpha = 0.1$			$\alpha = 0.2$			$\alpha = 0.3$			$\alpha = 0.4$			$\alpha = 0.5$		
			SS	DTC	SC	SS	DTC	SC	SS	DTC	SC	SS	DTC	SC	SS	DTC	SC
Location	APS	Stand	7.12	0.05	0.80	6.07	0.10	0.79	5.58	0.16	0.69	4.61	0.20	0.76	4.20	0.27	0.61
		KRPS	5.90	0.02	1.00	4.96	0.05	1.00	4.05	0.01	1.00	3.65	0.12	1.00	2.61	0.00	1.00
	RAPS	Stand	2.66	0.07	0.89	2.03	0.15	0.93	1.70	0.25	0.95	1.56	0.34	0.96	1.46	0.43	0.95
		KRPS	2.18	0.05	1.00	1.78	0.13	1.00	1.58	0.14	1.00	1.45	0.13	1.00	1.36	0.22	1.00
Action	APS	Stand	9.77	0.05	0.75	8.01	0.10	0.73	8.14	0.20	0.72	5.72	0.30	0.70	6.56	0.14	0.69
		KRPS	7.29	0.02	1.00	6.01	0.03	1.00	5.59	0.01	1.00	4.35	0.11	1.00	3.94	0.00	1.00
	RAPS	Stand	4.67	0.07	0.84	3.94	0.13	0.87	3.99	0.23	0.86	2.54	0.25	0.92	2.95	0.37	0.91
		KRPS	3.82	0.03	1.00	3.26	0.02	1.00	3.27	0.13	1.00	2.22	0.15	1.00	2.52	0.17	1.00

Table 8: Results on the ROAD dataset for the task sequences $\{agent \rightarrow location\}$ and $\{agent \rightarrow action\}$.

Task	Score	Method	$\alpha = 0.1$			$\alpha = 0.2$			$\alpha = 0.3$			$\alpha = 0.4$			$\alpha = 0.5$		
			SS	DTC	SC	SS	DTC	SC	SS	DTC	SC	SS	DTC	SC	SS	DTC	SC
Location	APS	Stand	9.02	0.07	0.79	7.92	0.10	0.76	5.97	0.18	0.73	4.03	0.22	0.71	3.98	0.29	0.64
		KRPS	7.50	0.01	1.00	5.99	0.06	1.00	4.00	0.01	1.00	3.24	0.14	1.00	2.73	0.05	1.00
	RAPS	Stand	2.45	0.09	0.87	2.07	0.15	0.91	1.87	0.23	0.93	1.83	0.34	0.96	1.40	0.40	0.95
		KRPS	2.01	0.04	1.00	2.04	0.10	1.00	1.35	0.14	1.00	1.17	0.09	1.00	1.10	0.09	1.00
Action	APS	Stand	10.41	0.05	0.77	6.84	0.13	0.74	6.02	0.15	0.74	5.36	0.12	0.78	4.83	0.20	0.68
		KRPS	5.98	0.02	1.00	4.18	0.02	1.00	3.90	0.07	1.00	3.87	0.12	1.00	3.94	0.08	1.00
	RAPS	Stand	4.36	0.06	0.87	3.03	0.15	0.89	2.48	0.16	0.86	1.50	0.24	0.94	1.20	0.37	0.90
		KRPS	3.57	0.04	1.00	2.88	0.02	1.00	2.03	0.09	1.00	1.07	0.07	1.00	1.02	0.10	1.00

Table 9: Results on the Waymo/ROAD++ dataset for the task sequences $\{agent \rightarrow location\}$ and $\{agent \rightarrow action\}$.

C Results for Further Task Sequences

In Section 4.5, we reported results for sequences of 2 tasks: $\{agent \rightarrow location\}$ and $\{agent \rightarrow action\}$ for $\alpha = [0.1, 0.2, 0.4]$. In the following, we present more results on both datasets for the full set of values of $\alpha = [0.1, 0.2, 0.3, 0.4, 0.5]$ for the sequences $\{agent \rightarrow location\}$ and $\{agent \rightarrow action\}$ in Table 8 and Table 9, respectively. Furthermore, we present results using sequences of 3 tasks on the ROAD dataset to show the capability of KRPS to handle sequences with higher numbers of tasks and in different task orders. The 3-tasks sequences that we consider are: $Seq1 : \{agent \rightarrow action \rightarrow location\}$, and $Seq2 : \{location \rightarrow action \rightarrow agent\}$. We use the same evaluation set-up and data splits reported in the evaluation section in our paper. The results are reported in Table 10 and Table 11, respectively.

In all task sequences, KRPS still holds the theoretical coverage guarantees for all values of α . More importantly, KRPS achieves the desired coverage rates while being able to reduce the set size considerably. The conditional semantic consistency also holds, as it is guaranteed by Theorem 3.1 and Corollary 3.1. For all task sequences, the conditional semantic consistency for task 2 with respect to task 1 and \mathcal{K} , task 3 with respect to task 2 and \mathcal{K} , and task 3 with respect to task 1 and \mathcal{K} , is 100%.

D Qualitative Results

In this section, we present further qualitative results on the ROAD and Waymo/ROAD++ datasets for the task sequences $\{agent \rightarrow action\}$, $\{agent \rightarrow location\}$, and $\{agent \rightarrow action \rightarrow location\}$.

Task	Score	Method	$\alpha = 0.1$			$\alpha = 0.2$			$\alpha = 0.3$			$\alpha = 0.4$			$\alpha = 0.5$		
			SS	DTC	SC	SS	DTC	SC	SS	DTC	SC	SS	DTC	SC	SS	DTC	SC
Action	APS	Stand	9.82	0.05	0.71	8.14	0.10	0.65	6.91	0.15	0.61	5.87	0.21	0.58	5.10	0.26	0.56
		KRPS	7.12	0.01	1.00	5.84	0.02	1.00	4.79	0.00	1.00	3.88	0.01	1.00	3.17	0.00	1.00
	RAPS	Stand	4.73	0.07	0.84	3.98	0.13	0.87	3.16	0.18	0.9	2.61	0.26	0.92	2.20	0.33	0.93
		KRPS	3.84	0.06	1.00	3.26	0.12	1.00	2.65	0.17	1.00	2.24	0.15	1.00	1.93	0.21	1.00
Location	APS	Stand	7.67	0.05	0.83	6.77	0.14	0.82	6.19	0.21	0.80	5.80	0.22	0.78	5.31	0.24	0.77
		KRPS	6.33	0.01	1.00	5.30	0.00	1.00	4.53	0.08	1.00	3.86	0.00	1.00	3.23	0.01	1.00
	RAPS	Stand	3.03	0.07	0.90	2.38	0.16	0.94	2.04	0.25	0.95	1.83	0.35	0.96	1.69	0.35	0.96
		KRPS	2.06	0.07	1.00	2.12	0.16	1.00	1.80	0.04	1.00	1.69	0.13	1.00	1.57	0.13	1.00

Table 10: Results on the ROAD dataset for the task sequence $\{agent \rightarrow action \rightarrow location\}$.

Task	Score	Method	$\alpha = 0.1$			$\alpha = 0.2$			$\alpha = 0.3$			$\alpha = 0.4$			$\alpha = 0.5$		
			SS	DTC	SC	SS	DTC	SC	SS	DTC	SC	SS	DTC	SC	SS	DTC	SC
Action	APS	Stand	9.82	0.05	0.72	8.15	0.00	0.68	6.91	0.15	0.64	5.87	0.21	0.60	5.10	0.26	0.58
		KRPS	7.35	0.00	1.00	6.02	0.00	1.00	4.91	0.00	1.00	3.97	0.00	1.00	3.20	0.01	1.00
	RAPS	Stand	4.73	0.07	0.82	3.98	0.15	0.85	3.16	0.20	0.89	2.61	0.26	0.91	2.20	0.35	0.92
		KRPS	3.86	0.06	1.00	3.24	0.13	1.00	2.64	0.18	1.00	2.23	0.25	1.00	1.93	0.32	1.00
Agent	APS	Stand	7.00	0.06	0.53	6.52	0.13	0.54	6.16	0.20	0.55	5.76	0.27	0.56	5.48	0.33	0.55
		KRPS	3.45	0.00	1.00	3.07	0.00	1.00	2.74	0.00	1.00	2.36	0.00	1.00	2.03	0.01	1.00
	RAPS	Stand	1.55	0.08	0.92	1.31	0.08	0.92	1.22	0.27	0.97	1.16	0.27	0.97	1.13	0.26	0.95
		KRPS	1.31	0.08	1.00	1.19	0.10	1.00	1.14	0.17	1.00	1.11	0.27	1.00	1.08	0.26	1.00

Table 11: Results on the ROAD dataset for the task sequence $\{location \rightarrow action \rightarrow agent\}$.

D.1 Qualitative Results on the ROAD Dataset

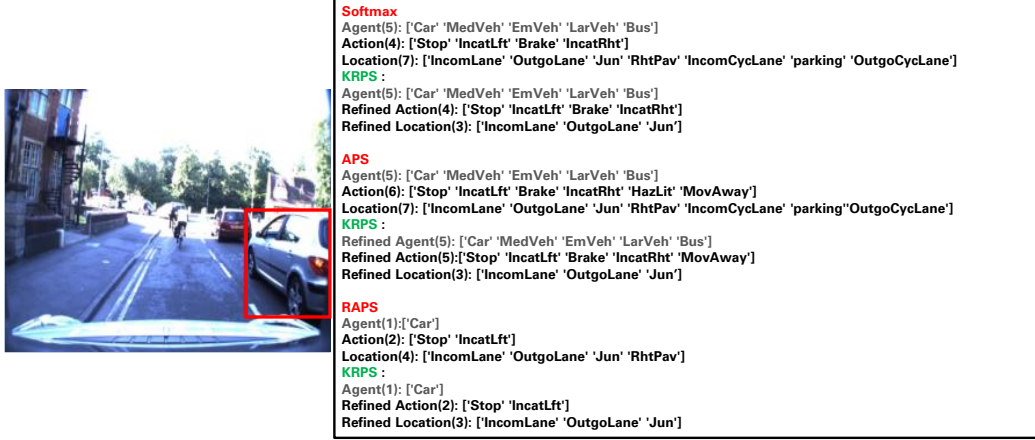
Figure 4 shows 3 different scenes from the ROAD dataset with the agent of interest highlighted with the red bounding box. For each bounding box, we perform the indicated CP procedure to acquire the prediction sets for the agent classification task. Based on the agent prediction sets, we report the generated prediction sets with and without KRPS using the Softmax, APS, and RAPS scores.

Figure 4a shows a scene with a *car stopping in the outgoing lane*. The prediction sets for the action and location classification tasks demonstrate how KRPS achieved a substantial reduction for the prediction sets. This reduction is particularly observable for the location task with the softmax and APS scores, where the prediction set size is reduced by 4 classes.

Figure 4b shows a scene with a *bus stopping in the vehicle lane*. All approaches succeed in including the correct labels in the prediction sets. The combination of KRPS with RAPS succeeds in constructing a singleton for the tasks location and action classification. By applying KRPS, the action and location classes that are not relevant to the agent class are removed.

Figure 4c showcases a scene with a *bus moving towards in the incoming lane*. Using KRPS, the size of the prediction sets for the action classification is reduced by 50% for the softmax score, by 75% for APS, and by 66% for RAPS. For the location classification task, KRPS reduced the prediction set size by 50% for the softmax score, by 50% for APS. For RAPS, the prediction set is not reduced, since both locations, *incoming lane* and *junction* are possible, given the agent class and its action.

The figures highlight the capability of KRPS to reduce the set size by removing action and location classes that are not relevant to the agent. Theorem 3.1 ensures that this removal procedure does not affect the marginal coverage property.



(a) Scene of a car stopping in the outgoing lane from the ROAD dataset [1] with prediction sets using 3 scoring functions for CP (Softmax, APS, RAPS) without and with KRPS.



(b) Scene of a bus stopping in the vehicle lane from the ROAD dataset [1] with prediction sets using 3 scoring functions for CP (Softmax, APS, RAPS) without and with KRPS.



(c) Scene of a bus moving towards in the incoming lane from the ROAD dataset [1] with prediction sets using 3 scoring functions for CP (Softmax, APS, RAPS) without and with KRPS.

Figure 4: Scenes from the ROAD dataset [1] with prediction sets using 3 scoring functions for CP (Softmax, APS, RAPS) without and with KRPS.

D.2 Qualitative Results on the Waymo/ROAD++ Dataset

Figure 5 shows 2 different scenes from the Waymo/ROAD++ dataset, characterized by challenging situational and environmental conditions that may induce high uncertainty. The agent of interest is highlighted with the red bounding box. We perform the indicated CP procedure for each bounding box to acquire the prediction sets for the agent classification task. Based on the agent prediction sets, we report the generated prediction sets with and without KRPS using the Softmax, APS, and RAPS scores.

Figure 5a depicts a low-light scenario where a pedestrian crosses the street with a car in the background of the bounding box. The challenging lighting conditions and complex scene composition contribute to uncertainty, prompting the model to assign vehicle-associated actions and locations such as *Brake* and *outgoing lane*. using KRPS mitigates this confusion by restricting the subsequent tasks to consider only classes suitable for pedestrians or bicycles, as determined by the agent classification model. KRPS notably reduces uncertainty and shrinks the prediction set size by 80%, 83%, and 50% for softmax, APS, and RAPS predictions, respectively, for the action classification task. For location classification, the prediction set size is reduced by 50%, 75%, and 50% for softmax, APS, and RAPS, respectively.

Figure 5b illustrates a scenario where a pedestrian is crossing the street while pushing a bicycle. This presence of the bicycle often misleads models for action and location to assign characteristics typical of bicyclists, such as *Brake* and *outgoing lane*. KRPS addresses this issue by ensuring that only classes suitable for either pedestrians or bicycles are considered, as dictated by the initial agent classification results. This application of KRPS significantly reduces uncertainty and narrows the prediction set size by 71%, 60%, and 66% for the softmax, APS, and RAPS predictions for the action classification task, respectively. Similarly, for the location classification task, the prediction set sizes are reduced by 66%, 50%, and 66% for softmax, APS, and RAPS, respectively.



(a) Scene of a *pedestrian pushing a bicycle and crossing the street* from the Waymo/ROAD++ dataset with prediction sets using 3 scoring functions for CP (Softmax, APS, RAPS) without and with KRPS.



(b) Scene of a *pedestrian pushing a bicycle and crossing the street* from the Waymo/ROAD++ dataset with prediction sets using 3 scoring functions for CP (Softmax, APS, RAPS) without and with KRPS.

Figure 5: Scenes from the Waymo/ROAD++ dataset with prediction sets using 3 scoring functions for CP (Softmax, APS, RAPS) without and with KRPS.