From Objects to Events: Unlocking Complex Visual Understanding in Object Detectors via LLM-guided Symbolic Reasoning

Supplementary Material

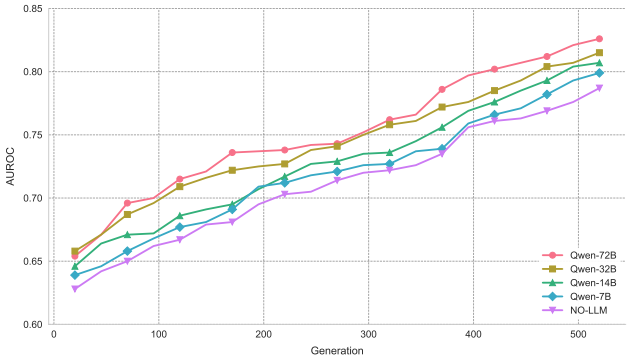

Figure 1. Performance on SymbolicDet with or without LLM.

	Ours	NSCL	NS-VQA	LogicHOI
Multi-rods	0.75	0.65	0.61	0.67
Helmet	0.83	0.68	0.65	0.69

Table 1. Comparison with other nuero-symbolic methods. (Acc)

1. Ablation Study

Analysis of Component Contributions. To understand the contribution of each component in our framework, we conduct comprehensive ablation studies examining the individual and combined effects of LLM reasoning and symbolic regression. Starting with a baseline using only manual logic expressions (67.00% average performance), the addition of symbolic regression significantly improves performance to 85.36%. This substantial improvement (+18.36%) suggests that automated pattern discovery through symbolic regression is significantly more effective than human-designed rules, likely due to its ability to explore a broader space of logical combinations and capture subtle patterns that might not be immediately apparent to human experts.

Impact of LLM Integration. Figure 1 illustrates the substantial impact of different LLM integration on both the effectiveness and efficiency of our symbolic pattern discovery process. When examining convergence trajectories across generations, we observe that LLM guidance not only enhances the ultimate detection accuracy but also significantly accelerates the convergence speed of symbolic regression. The analysis compares performance curves with and without LLM guidance, as well as across different LLM scales. *Finding: Effective event detection through symbolic reasoning benefits from the complementary strengths of systematic pattern discovery (through evolutionary search) and semantic guidance (through LLM reasoning). The symbolic component provides the expressive framework for capturing complex relationships, while the LLM component contributes domain knowledge and conceptual understanding that steers the search toward meaningful patterns.*

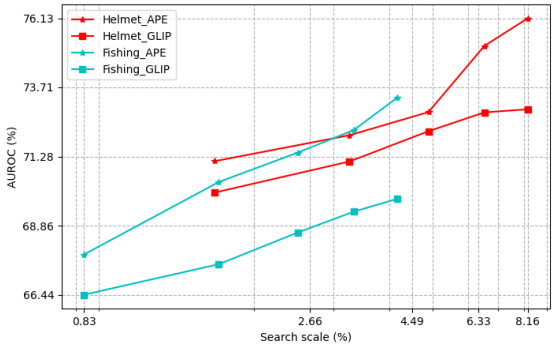

Figure 2. Performance on different search scales.

	wo llm	7B	14B	32B	72B
Run time(s/500it)	80.66	265.5	281.01	268	267
Cost time(s)[1]	69.66	45.92	44.91	30.35	26.07
Memory(MB)	293.65	218.56	218.34	218.68	218.82

Table 2. The computational overhead of symbolic search process.
[1] It refers to the time needed to achieve the same performance.

Effect of Search Scale on Performance. To further explore the robustness of our framework, we investigate the effect of varying search scales on event detection accuracy, as depicted in Figure 2. The search scale defines the proportion of samples allocated for constructing the logical search space, with the remainder used for pattern evaluation. Our results reveal a clear pattern: increasing the search scale consistently enhances AUROC performance across both the Helmet and Fishing datasets using APE and GLIP strategies. Notably, in the Helmet dataset, both strategies show a significant improvement, reaching peak performance at the highest search scale of 8.16%. The Fishing dataset demonstrates a similar upward trend, highlighting the benefits of expanding the search space. *Finding:* The increase in performance with larger search scales underscores the efficacy of our approach in utilizing more extensive logical reasoning. The findings suggest that even without traditional fine-tuning, enlarging the search space enables the framework to uncover more accurate and interpretable patterns. This scal-

	Detection	Search	LLm Infer
Run time(s/iter)	0.02-1.5	0.16	8-14
GPU Memory(GB)	0.54-18	-	-

Table 3. The computational overhead of each part.

	SPORT	CONCERT	PROTEST
Ours	0.93	0.99	0.92
USED	0.66	0.75	0.67

Table 4. The computational overhead of each part. (Acc)

ability evidences the flexibility and potency of SymbolicDet in capitalizing on the latent potential of standard object detectors, reinforcing its applicability across diverse scenarios.

2. Extra experiments

We conducted additional experiments by transferring several representative neuro-symbolic methods to be evaluated on our benchmark dataset. Detailed results can be found in Table 1. In addition, we conducted experiments to compare the computational overhead. We analyze the trade-off between computational cost and performance when using LLMs with varying parameter counts versus not using an LLM at all.Details can be found in Table 2. We also analyze the computational costs of each step of our work.The main computational steps of this work are: Detector inference, Symbolic search process and LLM-guided reasoning. It is worth noting that we use a third-party LLM service provider for the third step of computation, so the actual computational cost may vary significantly depending on the service provider. Detailed computational costs for each part are shown in Table 3.
In addition, we have conducted additional experiments on three subsets of the USED dataset: SPORT, CONCERT, and PROTEST. The results from these new experiments will be presented in Table 4.

3. Notation and Results

This section provides a comprehensive list of mathematical symbols and notations used throughout this paper, followed by visual demonstration of our proposed method's performance. Table 5 summarizes the key mathematical symbols and their definitions used in this work. Figure 3 illustrates the effectiveness of our proposed approach through visual comparison and performance metrics.

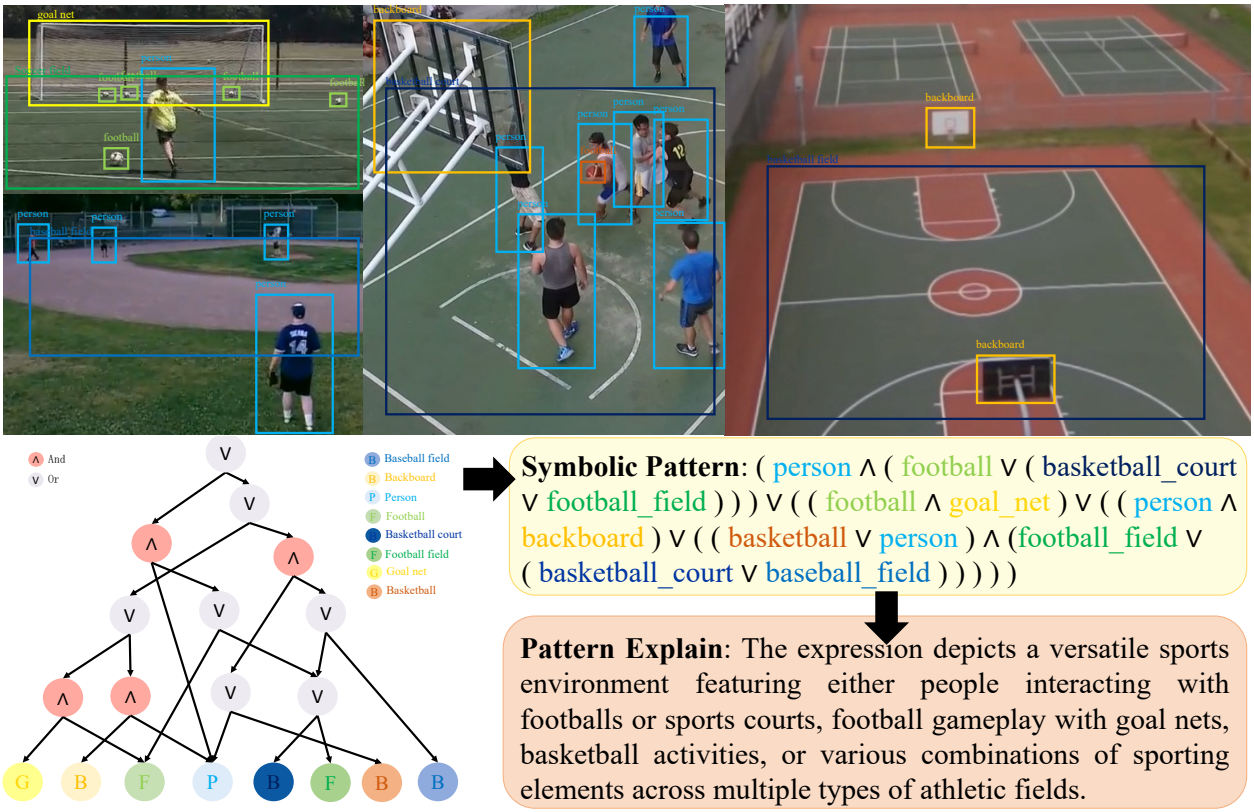

Figure 3. Illustration of the application of symbolic pattern detection in sports environments, showcasing how logical expressions can be used to identify and categorize complex sports scenarios.

Symbol	Definition
$\mathcal{D}$	A visual datsdet
I_i	An image i in visual dataset
y_i	A binary label corresponds to image i
D	A set of objects which are detected by object detector
d_j	A detected object
c_j	Category label belongs to d_j
b_j	Bounding box belongs to d_j
s_j	Confidence score belongs to d_j
ϵ	Target event
f	Symbolic expression
$\mathcal{O}_I$	A binary symbol to judge whether the target event is present within an image I
f^*	Optimal discovered symbolic expression
$\mathcal{F}$	The space of all possible expressions in our symbolic language
$\mathcal{T}$	Object detector
S	A scoring function that evaluates how well an expression distinguishes positive and negative examples
$\mathcal{G}_{LLM}$	The LLM guidance mechanism that directs the search toward promising expressions
$\phi_i(\cdot)$	Feature extraction functions that capture entity counts, spatial relationships, and attribute distributions
$\Omega(f)$	A complexity penalty that promotes simpler expressions
P_{init}	Initial prompt for the LLM guidence
P_{cot}	Prompt for chain-of-thought
P_{feed}	Prompt after contextual feedback integration

Table 5. Symbol Definition