NEURAL INFORMATION PROCESSING SYSTEMS

SEOULTECH

Are Multiple Instance Learning Algorithms Learnable for Instances?

Jaeseok Jang and HYUK-YOON KWON*

Big Data-Driven AI Laboratory (https://bigdata.seoultech.ac.kr/)

Graduate School of Data Science, Seoul National University of Science and Technology
{jangjs1027, hyukyoon.kwon}@seoultech.ac.kr

I. Introduction

- **Multiple Instance Learning (MIL)**
  - **Learns from labels assigned to bags and performs predictions at both the bag and instance levels.**
    - **Ex) Video-Snippets, Review-Words, Image-Patch, Sliding window-Time point**
  - **Most research primarily focuses on enhancing prediction performance at the bag level, rather than the instance level.**

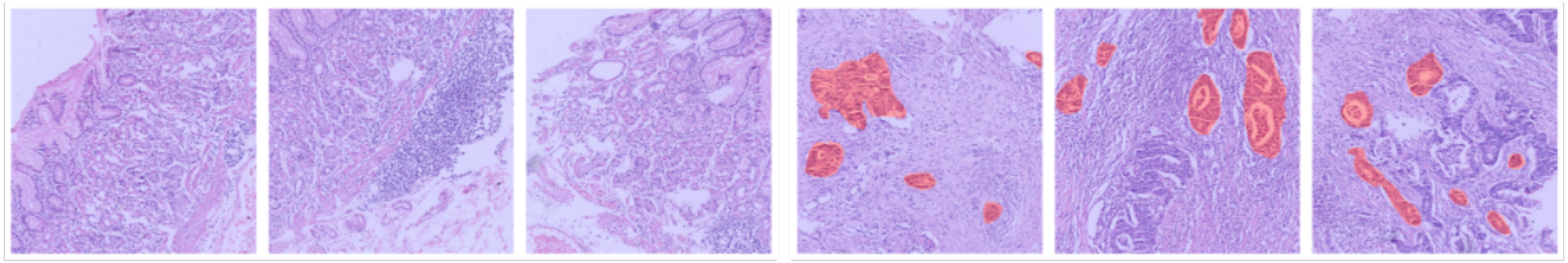


Figure 1. The data structure consisting of multi-instances (Blue: Negative, Red: Positive)

II. Problem Setting

- **Definition 4. PAC Learnability of Bag**

$$\mathbb{P}_{S\sim D_{XY}^m}[\,|R_{bag}(A(S)) - \inf_{h\in\mathcal{H}_{bag}} R_{bag}(h)\,| \leq \epsilon] \geq 1-\delta$$

- **Definition 5. PAC Learnability of Instance**

$$\mathbb{P}_{S_{inst_i}\sim D_{X_{inst_i}Y}^m}[\,|R_{inst_i}(A(S_{inst_i})) - \inf_{h\in\mathcal{H}_{inst_i}} R_{inst_i}(h)\,| \leq \epsilon] \geq 1-\delta$$

- **Definition 6.** *If the MIL algorithm satisfies Condition 2, it is learnable for instances.*

  Condition 2. *The Deep MIL algorithm A must exhibit equivalent PAC learnability for bags and instances*:

$$\mathbb{P}\left[\,|R_{bag}(A(S)) - \inf_{h\in\mathcal{H}_{bag}} R_{bag}(h)\,| \leq \epsilon \wedge \bigcap_{i=1}^{n} |R_{insti}(A(S_{inst_i})) - \inf_{h\in\mathcal{H}_{inst_i}} R_{inst_i}(h)\,| \leq \epsilon\right] \geq 1-\delta$$

II. Problem Setting

- **Theorem 1.** *If MIL algorithm A satisfies Condition 1, then this algorithm is not PAC learnable for any instance domain space $\mathscr{D}_{X_{inst_i}Y}$ and instance hypothesis space $\mathscr{H}_{inst_i} \subset \{h_{inst_i} : X \to Y\}$.*

$$\mathbb{P}\left[\bigcup_{i=1}^{n} |R_{inst_i}(A(S_{inst_i})) - \inf_{h \in \mathscr{H}_{inst_i}} R_{inst_i}(h)| > \epsilon\right] > \delta$$

- **Condition 1.** *The MIL algorithm A is not PAC learnable for the given domain space $\mathscr{D}_{XY}$ and bag hypothesis space $\mathscr{H}_{bag} \subset \{h_{bag} : X \to Y\}$:*

$$\mathbb{P}\left[|R_{bag}(A(S)) - \inf_{h \in \mathscr{H}_{bag}} R_{bag}(h)| > \epsilon\right] > \delta$$

- **Assumption 1.** *The MIL algorithm is PAC learnable for bags.*

III. Proposed Theoretical Framework

III.I. Summary of Framework

Priori Learnability Theorems for Instances

Theorem 1: Relationship between learnability for Bags and learnability for Instances
Consequence: If an MIL algorithm is not PAC learnable for bags, it cannot be PAC learnable for instances.

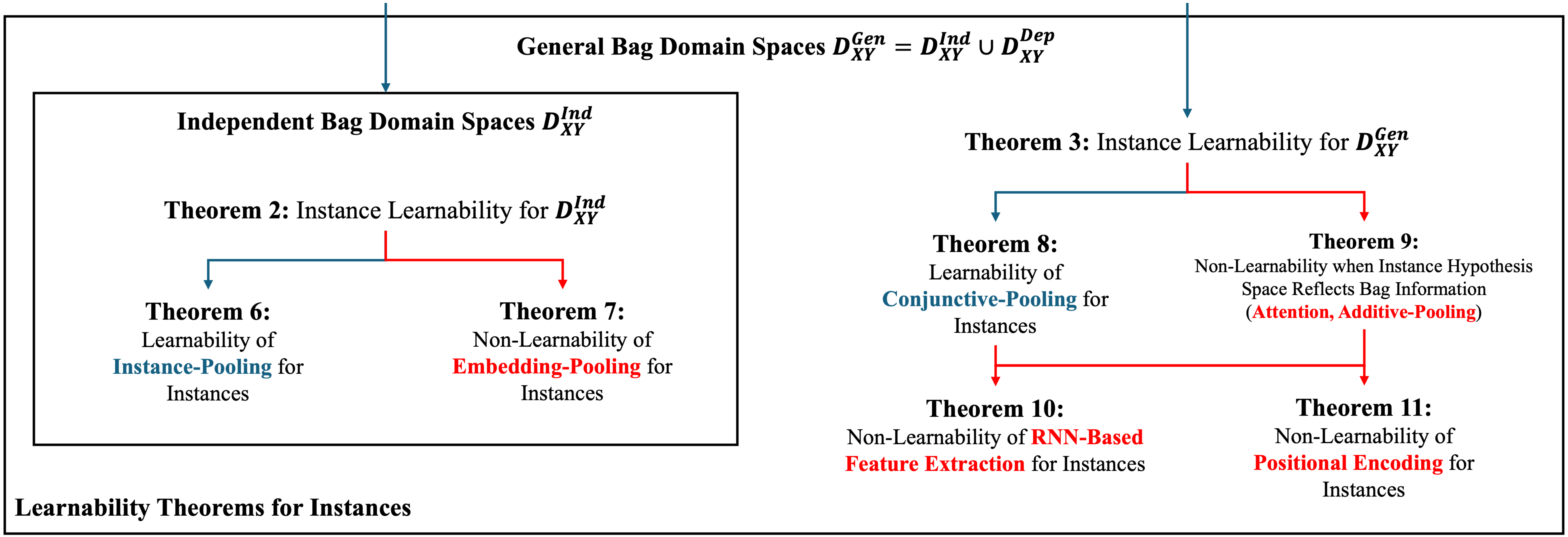
Assumption 1: The MIL algorithm is PAC learnable for bags

Theorem 4: PAC Learnability for Bags in $\boldsymbol{D_{XY}^{Ind}}$
Consequence: All MIL pooling techniques are PAC trainable on D_{XY}^{Ind} to Bag.
(**Instance, Embedding, Attention, Additive, Conjunctive-Pooling**)
Example: In **drug activity prediction**, MIL analyzes various independent chemical structures of a drug candidate to predict whether the candidate is active against a disease.

Theorem 5: PAC Learnability for Bags in $\boldsymbol{D_{XY}^{Gen}}$
Consequence: For MIL algorithms to be PAC learnable in D_{XY}^{Gen}, the use of an attention mechanism is necessary.
(**Attention, Additive, Conjunctive-Pooling**)
Example: In **review-sentence analysis**, MIL considers the relationships between sentences within a review to predict the overall sentiment of the review.

III. Proposed Theoretical Framework

III.I. Summary of Framework



III. Proposed Theoretical Framework

III.II. PAC Learnability for Independent Bag Domain Spaces

- **Definition 7. Independent Bag Domain Space ($\mathscr{D}_{XY}^{Ind}$)**

$$D_{XY}^{Ind} := \bigcup_{i=1}^{N} D_{X_{inst_i}Y} \in \mathscr{D}_{XY}^{Ind}$$

- **Theorem 2.** *If a MIL algorithm satisfies Condition 4 in $\mathscr{D}_{XY}^{Ind}$, it is learnable for instances.*

  Condition 4. *The risk of the optimal hypothesis for D_{XY}^{Ind} must ensure that it equals the sum of the individual risks of the optimal hypotheses within D_{XY}^{Ind}:*

$$\inf_{h \in \mathscr{H}} R_{\mathscr{D}_{XY}^{Ind}} = \sum_{i=1}^{N} \inf R_{inst_i}$$

III. Proposed Theoretical Framework

III.III. PAC Learnability for General Bag Domain Spaces

- **Definition 8. General Bag Domain Space ($\mathscr{D}_{XY}^{Ind} \cup \mathscr{D}_{XY}^{Dep} = \mathscr{D}_{XY}^{Gen}$)**

$$D_{XY}^{Gen} = \sum_{i=1}^{N} \alpha_i D_{X_{inst_i}Y} \in \mathscr{D}_{XY}^{Gen} \quad \textbf{such that} \quad \sum_{i=1}^{N} \alpha_i = 1, \quad 0 \leq \alpha_i \leq 1$$

- **Theorem 3.** *If a MIL algorithm satisfies Condition 7 in $\mathscr{D}_{XY}^{Gen}$, it is learnable for instances.*

  Condition 7. *The risk of the optimal hypothesis for D_{XY}^{Gen} must ensure that it equals the weighted sum of the individual risks of the optimal hypotheses within D_{XY}^{Gen}:*

$$\inf_{h \in \mathscr{H}} R_{\mathscr{D}_{XY}^{Gen}} = \sum_{i=1}^{N} \alpha_i \inf R_{inst_i} \quad \textit{such that} \quad \sum_{i=1}^{N} \alpha_i = 1, \quad 0 \leq \alpha_i \leq 1$$

IV. Theoretical Verification of Existing Deep MILs

IV.I. Classifications of Existing Deep MIL Methodologies

- **Aggregation-level**
  - **At which stage are the values of individual instances aggregated?**
- **Attention-Target**
  - **At which stage are attention weights applied to the instances?**

Table 4: Classification of existing Deep MIL methodologies

	Instance -Pooling	**Embedding -Pooling**	**Attention -Pooling**	**Additive -Pooling**	**Conjunctive -Pooling**
Aggregation-level	Instance	Embedding	Embedding	Instance	Instance
Attention-target	None	None	Embedding	Embedding	Instance

IV. Theoretical Verification of Existing Deep MILs

IV.II. Verification Learnability for Instances by MIL Pooling Method

- **Lemma 1.** *Condition 9 serves as a necessary condition for the learnability of instances, when the hypothesis space for the i^{th} instance of a MIL algorithm is* $\mathscr{H}_{inst_i} \cup \mathscr{H}_{add_i}$*.*

    - **Condition 9.** $\mathscr{H}_{add_i}$ *must be a subset of* $\mathscr{H}_{inst_i}$*:*

$$\mathscr{H}_{inst_i} \supset \mathscr{H}_{add_i} := \{h_{add_i} : \mathscr{X}_{add_i} \rightarrow \mathscr{Y}\}$$

- **Theorem 6.** *In* $\mathscr{D}^{\mathbf{Ind}}_{\mathbf{XY}}$*, MIL algorithms that perform* *<u>instance-pooling</u>* *are* *<u>PAC learnable</u>* *for instances.*

- **Theorem 7.** *MIL algorithms that perform* *<u>Embedding-Pooling</u>* *are* *<u>not learnable</u>* *for instances.*

IV. Theoretical Verification of Existing Deep MILs

IV.II. Verification Learnability for Instances by MIL Pooling Method

- **Theorem 8.** *If the MIL algorithm does not adhere to Condition 10, it is not learnable for instances.*
  - **Condition 10.** *The risk R_{inst_i} for the i^{th} instance should be as follows:*

$$R_{inst_i} = \mathbb{E}_{(x_{inst_i}, y) \sim D_{X_{inst_i}Y}} \ell_{inst_i}(h, y) \quad \text{, where} \quad h \in \mathcal{H}_{inst_i} \cup \mathcal{H}_{bag-level_i}$$

  - Attention-Pooling and Additive-Pooling is not learnable for Instances
- **Theorem 9.** *When MIL algorithms use Conjunctive-Pooling for aggregation in $\mathcal{D}_{XY}^{Gen}$, they are learnable for instances.*

IV. Theoretical Verification of Existing Deep MILs

IV.II. Verification Learnability for Instances by MIL Pooling Method

Table 2: Prediction performance of Deep MIL on Bags in D_{XY}^{Gen}.

		Macro-F1 Score	Micro-F1 Score	Weighted-F1 Score
Instance-Pooling	mi-Net	0.3286	0.5548	0.4550
	Causal MIL	0.2341	0.3577	0.2645
	MIREL	0.3623	0.5318	0.4372
Attention-Pooling	Attention MIL	**0.7652**	**0.7683**	**0.7583**
	Loss-Attention	**0.7935**	**0.7832**	**0.7753**
	SA-AbMILP	**0.7540**	**0.7619**	**0.7562**
	TransMIL	**0.7834**	**0.7711**	**0.7738**
Additive-Pooling	Additive MIL	**0.5314**	**0.6341**	**0.5732**
Conjunctive-Pooling	Conjunctive MIL	**0.7544**	**0.7701**	**0.7683**
None-Pooling	Fully Connected Layer	**0.7704**	**0.7724**	**0.7714**

IV. Theoretical Verification of Existing Deep MILs

IV.II. Verification Learnability for Instances by MIL Pooling Method

Table 3: Prediction performance comparison of MIL algorithms on bags and instances.

	Performance of Bags (PB)		Performance of Instances (PI)		**PI - PB**	
	Macro-F1	AUROC	Macro-F1	AUROC	Macro-F1	AUROC
Attention MIL	0.8434	0.9516	0.3215	0.7317	-0.5219	-0.2199
Loss-Attention	0.8228	0.9574	0.4797	0.7951	-0.3431	-0.1623
SA-AbMILP	0.7692	0.9552	0.3340	0.5464	-0.4352	-0.4088
TransMIL	0.8515	0.9622	0.2192	0.5369	-0.6323	-0.4253
Additive MIL	0.4776	0.9181	0.2320	0.8092	-0.2456	-0.1089
Conjunctive MIL	0.7916	0.9463	0.6430	0.9516	**-0.1486**	**+0.0053**

IV. Theoretical Verification of Existing Deep MILs

IV.III.I. Rethinking Position Dependencies of Instances on Deep MILs

- **Theorem 10.** *If the MIL algorithm extracts features of instances through RNN-based neural networks for aggregation, it is unable to learn from instances.*
- **Theorem 11.** *If the hypothesis space $\mathscr{H}_{Pos-Encode_i}$ generated through positional encoding values for the i-th position of the MIL algorithm is not a subset of $\mathscr{H}_{inst_i}$, then the algorithm is not PAC learnable for instances.*

Table 4: Test positional dependencies for WebTraffic datasets

	Default	PE (All)	**PE (Att)**	PE (Predict)	RNN (All)	**RNN (Att)**	RNN (Predict)
AOPCR	13.041	12.372	14.555	12.256	9.011	**17.502**	12.210
NDCG@n	0.676	0.665	**0.727**	0.642	0.620	0.714	0.523

IV. Theoretical Verification of Existing Deep MILs

IV.III.II. Instance Learnability for Multi-dimensional Deep MILs

- **Multi-Dimensional(MD) MIL predicts multi-dimensional instances using a top-level bag label.**
- **MD-instances should consider relationships with other dimensions.**
    - **Conjunctive-Pooling, reflecting MD relationships through attention, showed the best performance.**

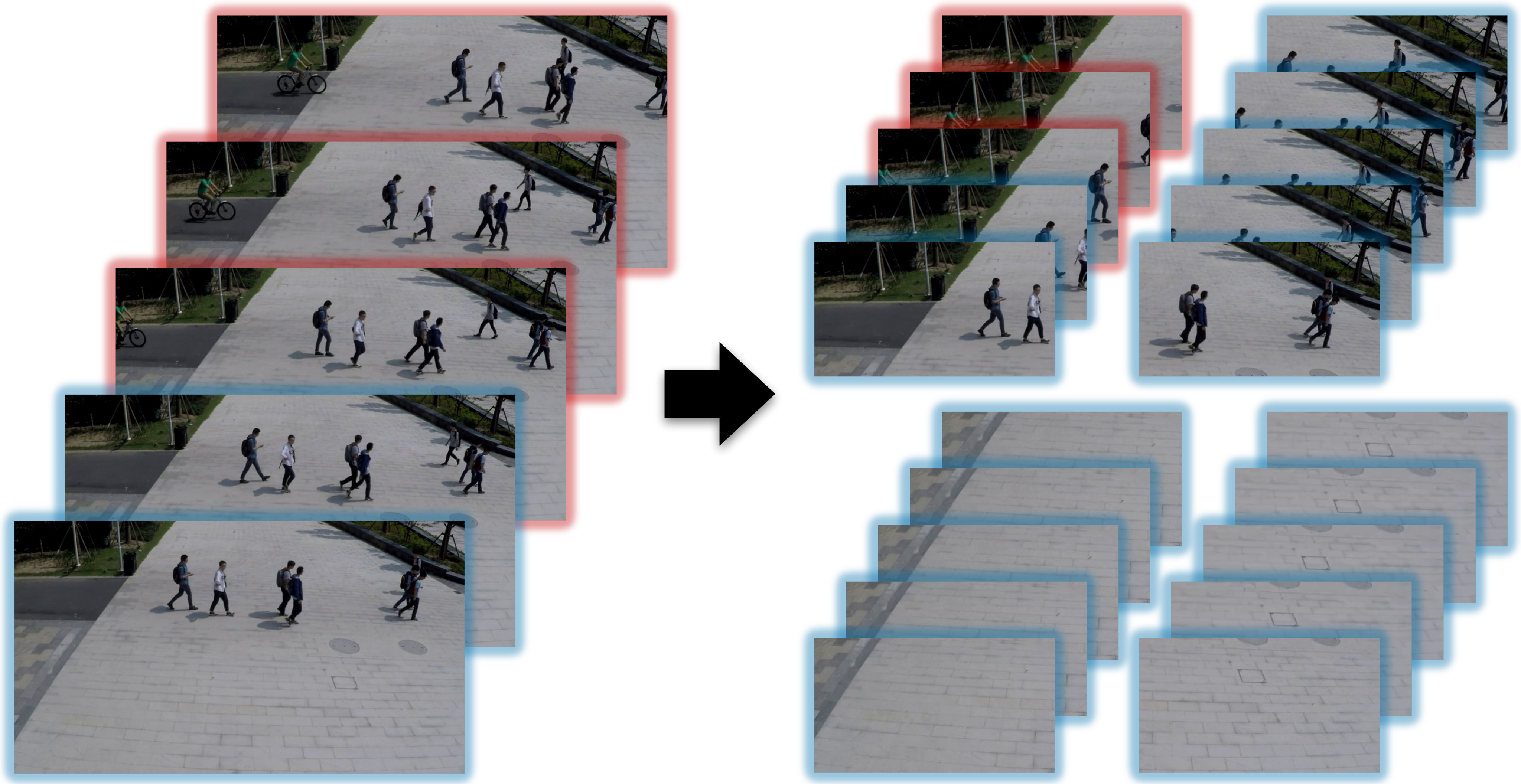


Table 4: Predicted performance for Snippets (i.e., bags) and patches (i.e., instances) of MD-MIL.

	None-Attention	Attention	**Cross-Attention**
Snippet (Bag)	0.87	0.88	**0.91**
Patch (Instance)	0.85	0.85	**0.91**

V. Conclusions

- **This study proposes a theoretical framework that defines the necessary conditions for an MIL algorithm to achieve learnability at the instance level, assuming Assumption 1 is satisfied.**
- **The framework is expected to benefit various domains where instance-level learnability is critical.**
    - **Although MIL is actively utilized in domains with limited labeling, such as medical applications, most research has focused on bag-level performance, primarily relying on Attention-Pooling methods.**
- **Future theoretical and experimental validations regarding positional dependencies and MD-MIL are anticipated to support further advancements in MIL research.**

Thank You