

Spatial-Logic-Aware Weakly Supervised Learning for Flood Mapping on Earth Imagery

Zelin Xu¹, Tingsong Xiao¹, Wenchong He¹, Yu Wang¹, Zhe Jiang^{1*},
Shigang Chen¹, Yiqun Xie², Xiaowei Jia³, Da Yan⁴, Yang Zhou⁵

¹University of Florida, Gainesville, FL, USA

²The University of Maryland, College Park, MD, USA

³University of Pittsburgh, Pittsburgh, PA, USA

⁴Indiana University Bloomington, Bloomington, IN, USA

⁵Auburn University, Auburn, AL, USA

{zelin.xu, xiaotingsong, whe2, yuwang1, zhe.jiang, sgchen}@ufl.edu,
xie@umd.edu, xiaowei@pitt.edu, yanda@iu.edu, yangzhou@auburn.edu

Abstract

Flood mapping on Earth imagery is crucial for disaster management, but its efficacy is hampered by the lack of high-quality training labels. Given high-resolution Earth imagery with coarse and noisy training labels, a base deep neural network model, and a spatial knowledge base with label constraints, our problem is to infer the true high-resolution labels while training neural network parameters. Traditional methods are largely based on specific physical properties and thus fall short of capturing the rich domain constraints expressed by symbolic logic. Neural-symbolic models can capture rich domain knowledge, but existing methods do not address the unique spatial challenges inherent in flood mapping on high-resolution imagery. To fill this gap, we propose a spatial-logic-aware weakly supervised learning framework. Our framework integrates symbolic spatial logic inference into probabilistic learning in a weakly supervised setting. To reduce the time costs of logic inference on vast high-resolution pixels, we propose a multi-resolution spatial reasoning algorithm to infer true labels while training neural network parameters. Evaluations of real-world flood datasets show that our model outperforms several baselines in prediction accuracy. The code is available at <https://github.com/spatialdatasciencegroup/SLWSL>.

Introduction

Flood extent mapping on high-resolution Earth imagery plays a crucial role in addressing major societal challenges, such as disaster management and response, national water forecasting, and energy and food security (Eftelioglu et al. 2017). For example, during Hurricane Harvey floods in 2017, first responders needed to know where flood water was to plan rescue efforts. In national water forecasting, detailed flood extent maps can be used to calibrate and validate the NOAA National Water Model, which can forecast the flow of over 2.7 million rivers and streams through the entire continental U.S. (Cline et al. 2009). Unfortunately, with the large amount of high-resolution earth imagery being collected from satellites (e.g., DigitalGlobe,

Spatial Knowledge Base for Flood Mapping

$r_1: \forall s_i, s_j \text{ Flood}(s_i) \wedge \text{Adjacent}(s_i, s_j) \rightarrow \text{Flood}(s_j)$
 $r_2: \forall s_i, s_j \text{ River}(s_i) \wedge \text{Distance}(s_i, s_j) \leq d \rightarrow \text{Flood}(s_j)$
 $r_3: \forall s_i, s_j \text{ Flood}(s_i) \wedge \text{River}(s_i) \wedge \text{Downstream}(s_i, s_j) \rightarrow \text{Flood}(s_j)$
 $r_4: \forall s_i \text{ Slope}(s_i) > s \rightarrow \neg \text{Flood}(s_i)$
 $r_5: \forall s_i \text{ Elevation}(s_i) > e \rightarrow \neg \text{Flood}(s_i)$



Figure 1: A real-world problem example.

Planet Labs), aerial planes (e.g., NOAA National Geodetic Survey), and unmanned aerial vehicles, the cost of manually labeling flood extent becomes prohibitive. Instead, there are abundant weak (imperfect) labels that are spatially coarse and noisy, e.g., data product from low-resolution Earth imagery, and non-expert annotations.

Given high-resolution Earth imagery with coarse and noisy training labels, a base deep neural network model, and a spatial knowledge base with label logic constraints, our problem is to infer the true labels in a high resolution while training neural network parameters. Figure 1 provides an example. The input high-resolution Earth imagery and coarse and noisy labels are shown on the left and right sides, respectively. The spatial domain logic rules on pixel labels are shown at the top. The goal is to infer the hidden true labels in a high resolution in the middle.

However, the problem poses several technical challenges. First, the spatially coarse and noisy labels make it hard to directly train a neural network on high-resolution Earth imagery. Second, hidden true pixel labels follow the physical constraints expressed by the spatial logic rules, requiring the learning framework to integrate data-driven and logic-guided approaches. Third, spatial uncertainty is inherent in the label inference process, coming from weak input labels, imperfect spatial logic rules, and the neural network training process. Finally, the computational costs are very high for spatial learning on a vast number of raster pixels, especially

*Corresponding author.

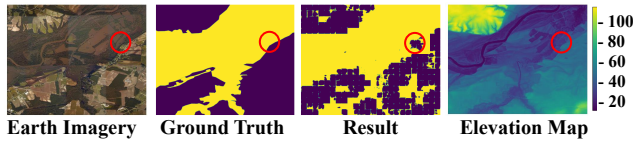


Figure 2: Physically implausible results.

when incorporating domain logic.

Numerous works exist that address weak (e.g., coarse) labels in earth imagery segmentation for convolutional neural networks or vision transformers. Techniques include utilizing teacher-student framework (Wang et al. 2020a; Guo et al. 2023), designing robust loss function (Mnih and Hinton 2012; Malkin et al. 2018), incorporating geometric properties of spatial labels into learning framework (He et al. 2022a; Jiang et al. 2022), and learning multi-scale features (Yang et al. 2012; Robinson et al. 2019; Cao and Huang 2022). However, these methods do not incorporate physical knowledge and thus may produce physically implausible results, i.e., erroneous predictions that violate physical constraints. An example is illustrated in Figure 2, where some pixels in the red circle are misclassified as dry (in purple) although they have lower elevation than nearby predicted flood pixels (in yellow).

Integration of domain knowledge and achievement of physical consistency by teaching models about the governing physical rules of the Earth system can provide very strong theoretical constraints on top of the observational ones (Reichstein et al. 2019). A few studies integrate domain knowledge into machine learning models (Rußwurm and Korner 2017; Fabrizio, Farina, and De Maio 2006; He et al. 2022b; Zhou et al. 2022; Jiang et al. 2023; Liu et al. 2023), but they mostly rely on incorporating specific physical properties (e.g., distance, topology) as fixed constraints. Such designs, while effective to some extent, are unable to incorporate flexible domain constraints expressed by rich logical expressions. In recent years, neural-symbolic systems have emerged as a promising solution, fusing symbolic logical reasoning with the prowess of deep neural networks (Garcez et al. 2022; Hu et al. 2016; Diligenti, Gori, and Sacca 2017; Donadello, Serafini, and Garcez 2017; Xu et al. 2018; Xie et al. 2019; Zhou et al. 2021; Cai et al. 2022). Several neural-symbolic models have been developed for weakly supervised learning based on pseudo-label generation (Manhaeve et al. 2018; Weber et al. 2019; Zhou 2019; Dai et al. 2019; Li et al. 2020; Tian et al. 2022; Duan et al. 2022). Nevertheless, these existing methodologies do not address the unique challenges posed by spatial data in flood mapping, especially spatial uncertainty and the high computational costs due to logical inference over vast pixels. Recently, a method (Xu et al. 2023) that utilizes a neural-symbolic framework for flood mapping was proposed. However, it is not designed for weakly supervised learning with coarse labels and fails to model label errors. Additionally, the model’s pseudo-label inference relies solely on spatial logic, neglecting the importance of image features.

To fill the gap, we propose a Spatial-Logic-aware Weakly

Supervised Learning framework (SLWSL) that combines deep learning and symbolic spatial logic to infer hidden true labels while training a deep learning model. We provide a probabilistic formulation of the overall sample distribution based on observed features, weak labels, and logical rules. Specifically, the unified likelihood objective is expressed by the fine-to-coarse label error model and a spatial-logic aware label probability model based on observed features and a knowledge base. The former functions as a prior probabilistic model capturing the error and noise of observed coarse labels, effectively simulating a weak annotator generating labels that deviate from the hidden ground truth. The latter characterizes both data-driven class-feature relationships and logical constraints. To solve this problem, we propose an expectation-maximization approach that iteratively infers hidden true labels and updates neural network parameters. To address the computational bottleneck of true label inference, we design a multi-resolution strategy to make a balance between spatial granularity and computational efficiency. The contributions of this paper are as follows:

- This paper proposes a novel spatial-logic-aware weakly supervised learning framework that integrates spatial domain logic into weakly supervised learning. We formulate a unified objective to characterize the weak label errors, the spatial label constraints, and the relationships between observed pixel features and hidden class labels.
- We develop a multi-resolution neural-symbolic learning algorithm to address the computational bottleneck in true label inference. Our idea is to use spatial uncertainty as a greedy heuristic to balance the trade-off between the spatial granularity of inferred labels and computational efficiency.
- Experiments on the real-world flood dataset during Hurricane Matthew in 2016 demonstrate the superior performance of SLWSL in classification over baseline methods.

Problem Statement

Preliminaries

Here, we introduce some concepts and notations for the spatial logic.

Definition 1. A *spatial raster framework* $\mathbf{S}$ is a two-dimensional grid composed of N cells. Each cell is considered a spatial data sample, denoted as $\mathbf{s}_i = (\mathbf{x}_i, y_i)$, with $1 \leq i \leq N$. It may include m non-spatial explanatory feature layers and a single class layer. The features of all samples are represented as $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, and the class is denoted by $\mathbf{Y} = \{y_1, y_2, \dots, y_N\}$, where $\mathbf{x}_i \in \mathbb{R}^{m \times 1}$ and y_i are the explanatory features and class of the i^{th} cell, respectively.

For instance, in flood mapping, the explanatory features may be spectral bands from remote sensing imagery, while the target classes represent flood and dry areas. Here, each image pixel is a spatial sample.

Definition 2. A *predicate* is a relation among objects or attributes of objects in the domain (e.g., *Adjacent*). An *atom*

is a specific instance of a predicate symbol applied to a tuple of terms (e.g., $\text{Adjacent}(s_i, s_j)$).

Definition 3. A **rule** is a clause constructed from atoms using logical connectives and quantifiers. An example could be: $\text{Flood}(s_i) \wedge \text{Adjacent}(s_i, s_j) \Rightarrow \text{Flood}(s_j)$.

Definition 4. A **ground atom** a and a **ground rule** r are specific variable instantiations of an atom and rule, respectively. The **grounding** of an atom or rule replaces all its arguments with constants. For example, for the rule: $\text{Flood}(s_i) \wedge \text{Adjacent}(s_i, s_j) \Rightarrow \text{Flood}(s_j)$, a ground rule should be: $\text{Flood}(s_1) \wedge \text{Adjacent}(s_1, s_2) \Rightarrow \text{Flood}(s_2)$, where s_1 and s_2 are specific spatial samples and $\text{Flood}(s_1)$, $\text{Adjacent}(s_1, s_2)$ and $\text{Flood}(s_2)$ are ground atoms.

With these definitions above, we can now formally establish the concept of a spatial knowledge base, $\mathcal{KB}$:

Definition 5. A **spatial knowledge base** $\mathcal{KB}$ is a collection of logic rules: $\mathcal{KB} = \{r_1, r_2, \dots, r_{|\mathcal{KB}|}\}$. Each rule r_i represents a spatial relationship, dependency, or constraint among entities within the set of spatial samples $\mathbf{S}$. $|\mathcal{KB}|$ denotes the number of rules in $\mathcal{KB}$.

While these rules are represented in first-order logic, they can be converted to a Markov Logic Network (Richardson and Domingos 2006) or Probabilistic Soft Logic framework (Kimmig et al. 2012; Bach et al. 2017), both of which blend probabilistic graphical models with first-order logic.

Problem Definition

Our problem can be formally defined as follows:

Input:

- A large-scale spatial raster framework comprised of N spatial samples denoted as $\mathbf{S} = \{s_1, s_2, \dots, s_N\}$.
- A set of explanatory features $\mathbf{X} = \{x_1, x_2, \dots, x_N\}$ within $\mathbf{S}$.
- A set of coarse and noisy labels $\tilde{\mathbf{Y}}$, where each label $\tilde{y}_i \in \{0, 1\}$.
- A spatial knowledge base $\mathcal{KB}$.
- A base neural network model, e.g., U-Net.

Output:

- Refined high-resolution labels, denoted as $\mathbf{Y}$.
- Trained deep learning model.

Objective:

- Maximize the prediction accuracy of the deep learning model.
- Maximize the performance of uncertainty estimation.

Constraints:

- The refined labels $\mathbf{Y}$ should be consistent with the spatial knowledge base $\mathcal{KB}$.

Our Proposed Approach

This section presents our proposed approach. We need to address several technical challenges: the mismatch between coarse and noisy labels and high-resolution image features, the difficulty in integrating domain knowledge (spatial logical rules) into the neural network learning process, and the high computational costs inherent to label inference

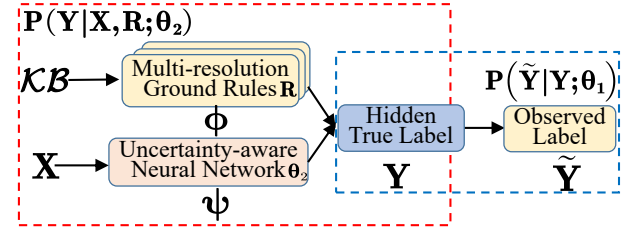


Figure 3: An overview of the overall framework.

for vast pixels. To address these challenges, we propose a spatial-logic-aware weakly supervised learning framework (SLWSL) that combines deep learning and symbolic spatial logic to infer hidden true labels while training a deep learning model. We provide a probabilistic formulation of the unified objective based on observed features, weak labels, and logical rules, which includes a label error model and a spatial-logic aware label probability model based on observed features and a knowledge base. We propose effective and efficient learning algorithms based on the expectation-maximization paradigm that iteratively infer hidden true labels and update neural network parameters. To address the computational bottleneck of true label inference, we design a multi-resolution strategy to make a balance between spatial granularity and computational efficiency.

Probabilistic Formulation of the Unified Objective

In this section, we introduce the overall objective of our SLWSL framework. Instead of directly using observed noisy label $\tilde{\mathbf{Y}}$ to train neural networks, we assume there exist hidden true labels $\mathbf{Y}$, influenced by both the features $\mathbf{X}$ and a set of logic rules $\mathbf{R}$. These rules are derived from the spatial knowledge base $\mathcal{KB}$ through a grounding process and describe the spatial relationships and dependencies. The overall log-likelihood of all samples can be expressed as follows:

$$\begin{aligned} \max_{\mathbf{Y}, \theta_1, \theta_2} L &= \log P(\tilde{\mathbf{Y}}, \mathbf{Y} | \mathbf{X}, \mathbf{R}; \theta_1, \theta_2) \\ &= \log P(\tilde{\mathbf{Y}} | \mathbf{Y}; \theta_1) + \log P(\mathbf{Y} | \mathbf{X}, \mathbf{R}; \theta_2) \end{aligned} \quad (1)$$

Here, the term $P(\tilde{\mathbf{Y}} | \mathbf{Y}; \theta_1)$ models the process of an annotator generating weak labels that deviate from the ground truth and serves as a probabilistic model for label errors (with parameters θ_1). $P(\mathbf{Y} | \mathbf{X}, \mathbf{R}; \theta_2)$ represents the likelihood of the predicted true labels given spatial rules and input feature (with parameters θ_2).

Label error model: In our probabilistic framework, $P(\tilde{\mathbf{Y}} | \mathbf{Y}; \theta_1)$, the prior probability of observing the coarse and noisy labels $\tilde{\mathbf{Y}}$ given the true labels $\mathbf{Y}$, encapsulates the imperfections and noise in the labeling process and serves as a bridge to connect the weak labels with hidden true labels. To model this prior probability, a key challenge is that the observed label $\tilde{\mathbf{Y}}$ and true label $\mathbf{Y}$ have different resolutions, i.e., one coarse observed label $\tilde{y}_i$ corresponds to a group of high-resolution labels $\{y_j\}$. Therefore, the label errors come from two parts: coarse resolution, and random

flipping noise. The coarse resolution can be expressed by a mapping from hidden true labels in a high-resolution to a coarse resolution, i.e., $y/i = \lfloor \frac{\sum y_i}{C^2} \rfloor$, where C is the resolution difference (e.g., one pixel in coarse resolution corresponds to $C \times C$ high-resolution pixels). We round the mean to the nearest integer (assuming a binary label 0 or 1). The random flipping noise can be formulated by a label class transition probability matrix Θ_1 . Each entry (p, q) of this transition matrix represents the probability of observing a noisy label p when the actual refined label is q . The matrix gives us a quantifiable measure of how likely certain label noise is to occur.

Likelihood of labels given features and rules: To model the likelihood of the label inference model $P(\mathbf{Y}|\mathbf{X}, \mathbf{R}; \Theta_2)$, we need to capture two primary factors:

- **Data-driven unary potential:** This term is related to the dependencies between the labels and the observed features. Formally, we represent it as $\psi(y_i, \mathbf{x}_i; \Theta_2)$, where Θ_2 are the parameters of the base deep learning model. This unary potential models the likelihood of a label y_i given the features $\mathbf{x}_i$, captured by a neural network.
- **Logic-driven interaction potential:** This term is related to the spatial relationships and dependencies between the labels that are dictated by the grounded logic rules $\mathbf{R}$. We represent this as $\phi(y_i, y_j; \mathbf{r}_m)$, where $\mathbf{r}_m \in \mathbf{R}$ is a grounded rule, and y_i, y_j are labels associated with that rule.

It is noted that following the set of Hinge-loss Markov Random Field in PSL (Bach et al. 2017), the potential functions are defined in such a way that the more probable configurations have smaller potential, rather than larger potential, which is the opposite of the typical convention for potential functions. The overall energy function here is then the sum of the data-driven unary potentials and the logic-driven interaction potentials, that is:

$$\begin{aligned} & -\log P(\mathbf{Y}|\mathbf{X}, \mathbf{R}; \Theta_2) \\ &= \sum_i \psi(y_i, \mathbf{x}_i; \Theta_2) + \sum_m \sum_{(i,j) \in \mathbf{r}_m} \phi(y_i, y_j; \mathbf{r}_m) \end{aligned} \quad (2)$$

We express the unary potential as $\sum_i \psi(y_i, \mathbf{x}_i; \Theta_2) = -\log P(\mathbf{Y}|\mathbf{X}, \Theta_2)$. This is the class likelihood of deep learning predictions.

To ensure consistency with spatial knowledge, we adopt t-norm fuzzy logic, mapping binary truth values to a continuous interval between $[0, 1]$. The logical conjunction ($\wedge$), disjunction ($\vee$), and negation ($\neg$) are defined as follows:

$$\begin{aligned} I(a_1 \wedge a_2) &= \max\{I(a_1) + I(a_2) - 1, 0\} \\ I(a_1 \vee a_2) &= \min\{I(a_1) + I(a_2), 1\} \\ I(\neg a_1) &= 1 - I(a_1) \end{aligned} \quad (3)$$

where I is the truth function and a_1, a_2 represent two atoms in a rule. It is natural to represent a label y using a logic atom, e.g., in flood mapping, the label y of a spatial sample s can be represented as the truth value of atom $Flood(s)$. With these definitions, we can model the extent to which a

Algorithm 1: Spatial-logic-aware Weakly Supervised Learning Algorithm

Input: weak labels $\tilde{\mathbf{Y}}$, features $\mathbf{X}$, knowledge base $\mathcal{KB}$

Parameters: maximum resolution level K , resolution constant η , error model Θ_1 , deep neural network Θ_2

Output: Inferred true labels $\mathbf{Y}$

- 1: Initialize Θ_1 and Θ_2 .
 - 2: Initialize resolution level $k \leftarrow K$.
 - 3: **while** $k > 0$ **do**
 - 4: Infer hidden true label $\mathbf{Y}$ with Θ_1 and Θ_2 fixed
 - 5: Update Θ_1 and Θ_2 with $\mathbf{Y}$ fixed
 - 6: $k \leftarrow k - 1$
 - 7: **end while**
 - 8: **return** $\mathbf{Y}, \Theta_1, \Theta_2$
-

rule is satisfied as a distance $d_r(I)$ to satisfaction for a rule $r : r_{body} \rightarrow r_{head}$:

$$d_r(I) = \max\{0, I(r_{body}) - I(r_{head})\} \quad (4)$$

Thus, the potential for spatial logic can be defined as:

$$\phi(y_i, y_j; \mathbf{r}_m) = \omega_m [\max\{l(y_i, y_j), 0\}]^{p_m} \quad (5)$$

where l is a linear function derived from Equation 3 and 4, ω_m is the weight for this rule, and $p_m \in \{1, 2\}$ provides a choice to square the potential.

Thus, by minimizing the energy function, we essentially aim to maximize this conditional probability, which aligns with our goal of refining the initial weak labels to better suit the data and the rules.

Effective and Efficient Learning Algorithms

Based on the unified objective, we need to optimize with respect to three variables: Θ_1, Θ_2 , and $\mathbf{Y}$. We employ an EM-like iterative approach to achieve this. In the ‘‘E-step’’, we infer the hidden labels $\mathbf{Y}$, assuming fixed Θ_1 and Θ_2 . Unlike vanilla EM, here we directly infer the most likely hidden true labels $\mathbf{Y}$ instead of computing an expectation. During the ‘‘M-step’’, with $\mathbf{Y}$ fixed, we update Θ_1 and Θ_2 to maximize the likelihood. The optimization employs gradient descent on the overall objectives. The overall algorithm is shown in Algorithm 1.

However, for large-scale spatial data (e.g., high-resolution Earth imagery), grounding the rules on all high-resolution pixels will lead to too many atoms, making the logic inference computationally intractable. Therefore, we need to strike a balance between granularity, computational efficiency, and prediction accuracy.

Multi-resolution logic reasoning: Here, we introduce how to formulate the logic-driven potential at multi-resolution. We aim to represent the spatial logic inference of sample labels within raster frameworks. This is achieved through two steps: spatial grounding and defining a potential function.

Spatial grounding in our case refers to replacing the variables in the spatial knowledge base rules with specific instances, i.e., pixels in earth imagery. To balance inference

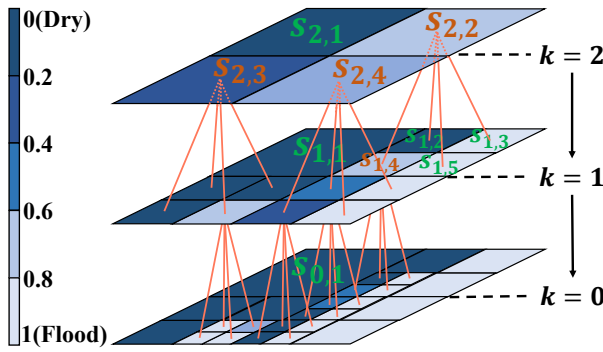


Figure 4: Illustration of hierarchical structure.

accuracy, efficiency, and granularity, we leverage a multi-resolution framework inspired by the fractal and hierarchical nature of spatial relationships.

We let the original input large-scale spatial raster framework, $\mathbf{S} = \mathbf{S}_0$, at the finest pixel-level resolution, and a constant η , where $\eta \in \mathbb{N}$ and $\eta > 1$, which helps us derive a set of resolutions: $1, \eta, \eta^2, \dots, \eta^K$ corresponding to the grid sizes at $K + 1$ levels. Here, the coarsest level has the same resolution of the initial weak label $\tilde{\mathbf{Y}}$, i.e., $\eta^K = \mathcal{C}$. As shown in Figure 4, these levels form a hierarchy with the coarsest resolution at the top (the root of the hierarchy) and the finest at the bottom. Each spatial sample $s_{k,i}$ at layer k corresponds to a group of cells at the next finer layer $k - 1$.

We propose a strategy that selects uncertain cells in a coarse layer to refine. The quantified uncertainty $u_{k,i}$ for each cell i at the k -th resolution level is calculated using the entropy of the inferred label $\hat{y}_{k,i}$ (He and Jiang 2023). Starting from the coarsest resolution ($k = K$), the process continues iteratively until the finest resolution ($k = 0$) is reached, selecting a subset of cells with the highest uncertainty at each resolution to refine the spatial partitioning.

Taking layer 2 in Figure 4 as an example, each cell in this grid is color-coded to denote the probability of dry (dark) and water (light). We view the cell with the leftmost and rightmost color (certain Dry and Flood) in the color bar as certain cells, and others as uncertain cells, i.e., only the $s_{2,1}$ can be viewed as a certain cell in layer 2. Therefore, the uncertain coarser cells in layer 2 are split into 2×2 finer cells, respectively.

Uncertainty-aware multi-instance learning: In this module, we model the data-driven potential to capture the information from the imagery features. Instead of binary labels, to directly optimize for the accuracy of the predicted probabilities, we work with a modified Binary Cross Entropy loss function where ground truth labels $y_{k,i}$ are replaced with uncertain labels $\hat{y}_{k,i}$.

Since a hierarchical structure is used for label inference, we face a multi-instance learning scenario (Maron and Lozano-Pérez 1997; Foulds and Frank 2010; Zhou 2004) when the spatial domain is partitioned into non-overlapping cells. Given different resolution levels, we calculate an aggregate probability output for each pixel instead of assigning a single label.

To account for cell-level labels, we define the uncertainty-aware multi-instance loss function as:

$$\log P(\mathbf{Y}|\mathbf{X}, \Theta_2) = \sum_i \hat{y}_{k,i} \log P_{k,i} + (1 - \hat{y}_{k,i}) \log(1 - P_{k,i}) \quad (6)$$

The probability output $P_{k,i}$ for each cell sample $s_{k,i}$ is the average of the predicted probabilities p_j for samples within the coarse cell $s_{k,i}$. This aggregate approach accounts for different granularity levels in the spatial domain, enhancing the model’s flexibility and adaptability to various spatial scales.

Experiments

Experiments Setup

Dataset description: We use two real-world flood mapping datasets collected from North Carolina during Hurricane Matthew in 2016. The explanatory features comprise the red, green, and blue bands within the aerial imagery obtained from the National Oceanic and Atmospheric Administration’s National Geodetic Survey¹. Digital elevation imagery was sourced from the University of North Carolina Libraries² for spatial logic rule construction. Each piece of data was subsequently resampled to a 2-meter by 2-meter resolution to standardize the information. We have 500 100-by-100 image patches in Dataset 1, among which 450 are for training and 50 are for testing (from different sub-regions); and 3360 patches in Dataset 2, among which 2856 are for training and 504 are for testing.

Candidate methods: In the experiments, we compare our proposed SLWSL with various methods: **Base:** The base deep learning model trained with observed weak labels. **U-CAM** (Wang et al. 2020b): This method uses image-level labels for training with class activation maps (CAM) for inference. **DeepProbLog (DPL)** (Manhaeve et al. 2018): A logic programming language that integrates deep learning with probabilistic logic programming. **Abductive Learning (ABL)** (Dai et al. 2019): A framework that combines both reasoning and learning by training a neural network model and using a logic reasoner to validate and revise these predictions. **w/o SR:** A simplified variant of our proposed SLWSL without selective refinement, which removes the selection of uncertain areas during the grounding process.

Evaluation metrics: We used precision, recall, F1 score, and Accuracy on the flood mapping class to evaluate the pixel-level classification performance.

Implementation details: When implementing SLWSL and baselines, we considered U-Net, a powerful deep learning model for image segmentation, as the Base model. We set the same architecture for the U-Net model in all methods with 5 downsample operations, 5 upsample operations, and a batch normalization within each convolutional layer. For other candidate methods and SLWSL, we use the same pre-trained U-Net to initialize the deep learning model in these frameworks.

¹<https://www.ngs.noaa.gov/>

²<https://www.lib.ncsu.edu/gis/elevation>

Method	Class	Dataset 1				Dataset 2			
		Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy
Base	Dry	0.70 $\pm$ 0.10	0.74 $\pm$ 0.06	0.71 $\pm$ 0.03	0.80 $\pm$ 0.02	0.80 $\pm$ 0.05	0.90 $\pm$ 0.03	0.84 $\pm$ 0.03	0.82 $\pm$ 0.03
	Flood	0.86 $\pm$ 0.06	0.84 $\pm$ 0.03	0.85 $\pm$ 0.02		0.87 $\pm$ 0.05	0.75 $\pm$ 0.04	0.80 $\pm$ 0.03	
U-CAM	Dry	0.75 $\pm$ 0.06	0.77 $\pm$ 0.07	0.75 $\pm$ 0.01	0.82 $\pm$ 0.02	0.83 $\pm$ 0.06	0.89 $\pm$ 0.02	0.86 $\pm$ 0.02	0.84 $\pm$ 0.02
	Flood	0.86 $\pm$ 0.06	0.86 $\pm$ 0.02	0.86 $\pm$ 0.02		0.85 $\pm$ 0.03	0.79 $\pm$ 0.06	0.82 $\pm$ 0.02	
DPL	Dry	0.82 $\pm$ 0.02	0.82 $\pm$ 0.01	0.82 $\pm$ 0.01	0.87 $\pm$ 0.00	0.92 $\pm$ 0.01	0.86 $\pm$ 0.01	0.89 $\pm$ 0.00	0.87 $\pm$ 0.00
	Flood	0.90 $\pm$ 0.01	0.90 $\pm$ 0.01	0.90 $\pm$ 0.00		0.79 $\pm$ 0.02	0.87 $\pm$ 0.02	0.83 $\pm$ 0.00	
ABL	Dry	0.89 $\pm$ 0.01	0.78 $\pm$ 0.01	0.83 $\pm$ 0.01	0.87 $\pm$ 0.00	0.87 $\pm$ 0.02	0.93 $\pm$ 0.01	0.90 $\pm$ 0.00	0.88 $\pm$ 0.00
	Flood	0.86 $\pm$ 0.01	0.94 $\pm$ 0.00	0.90 $\pm$ 0.00		0.90 $\pm$ 0.02	0.83 $\pm$ 0.02	0.86 $\pm$ 0.00	
w/o SR	Dry	0.95 $\pm$ 0.02	0.88 $\pm$ 0.02	0.92 $\pm$ 0.00	0.94 $\pm$ 0.00	0.91 $\pm$ 0.01	0.93 $\pm$ 0.01	0.92 $\pm$ 0.00	0.91 $\pm$ 0.00
	Flood	0.93 $\pm$ 0.02	0.97 $\pm$ 0.01	0.95 $\pm$ 0.00		0.89 $\pm$ 0.01	0.87 $\pm$ 0.01	0.88 $\pm$ 0.00	
SLWSL	Dry	0.91 $\pm$ 0.01	0.94 $\pm$ 0.00	0.93 $\pm$ 0.01	0.95 $\pm$ 0.00	0.90 $\pm$ 0.03	0.97 $\pm$ 0.01	0.93 $\pm$ 0.01	0.92 $\pm$ 0.01
	Flood	0.97 $\pm$ 0.00	0.95 $\pm$ 0.01	0.96 $\pm$ 0.00		0.96 $\pm$ 0.02	0.87 $\pm$ 0.04	0.91 $\pm$ 0.01	

Table 1: Comparison on pixel-level classification.

For the multi-resolution structure of SLWSL, we set the grid size constant $\eta = 10$ and $K = 2$ which means there are 3 layers with grid size 100×100 , 10×10 and 1×1 respectively. We construct the spatial knowledge base for the flood mapping task based on distance and topology relationships. For the distance relationship, we directly use the neighborhood pair to model. For the elevation, we adopt a Hidden Markov Tree model (Xie, Jiang, and Sainju 2018) which can model the topological relationship of each location based on the elevation.

All the experiments were conducted on an AMD EPYC 7742 64-core Processor CPU and an NVIDIA A100 GPU equipped with 80 GB of memory. We executed each model five times to obtain the average performance and its deviation.

Comparison on Pixel-level Classification

As demonstrated in Table 1, our SLWSL method consistently outperforms the candidate models. The Base model and U-CAM, predictably, struggle across both datasets. This can be attributed to the fact that coarse and weak labels often fail to provide adequate supervision, especially when training high-resolution data. The models DPL and ABL, which make efforts to embed domain knowledge into the learning process, show improved results over the Base model. This underlines the significance of integrating spatial knowledge into model training. Yet, they still fall short of SLWSL’s performance. This discrepancy can be primarily traced back to label inaccuracies, a consequence of relying on weak labels and the overwhelming presence of extensive logic rules in a large spatial area. The standout results of SLWSL across both datasets not only show its capability in effectively incorporating spatial domain knowledge with deep learning techniques but also highlight its unique strength in utilizing multi-resolution label inference. This becomes particularly crucial in scenarios dominated by weak training labels, showcasing SLWSL’s robustness and adaptability.

Moreover, Figure 5 presents a comparative visualization of various models’ performance on Dataset 2. It should be noted that the weak label image is of coarse resolution, with each pixel representing a 100 by 100 pixel area in comparison

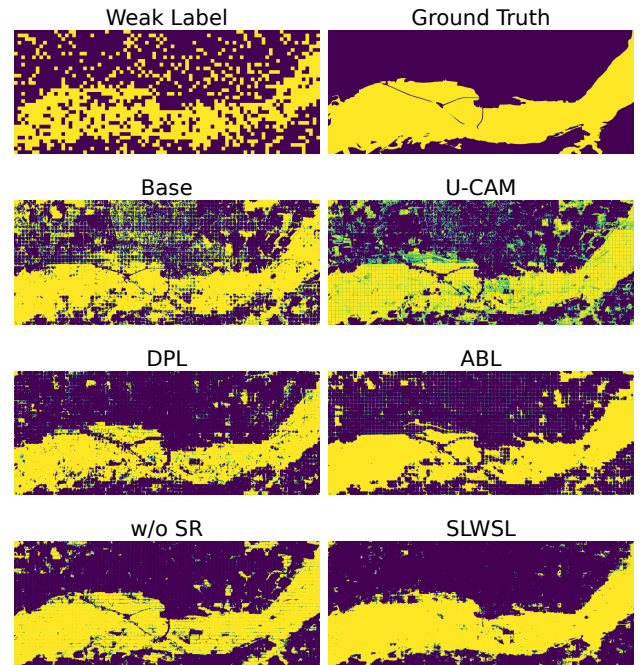


Figure 5: Visualization of classification results for dataset 2.

son to other images. Evidently, the Base and U-CAM models exhibit limitations in handling noise in the original label. Conversely, for SLWSL, a notable decrease in the number of misclassified pixels is clear, effectively reducing noise within each classified area.

Case Study

In this case study, the label inference efficacy of SLWSL is examined across various resolution levels during the training phase, as visually depicted in Figure 6. The figure highlights the evolution of inferred labels at three distinct resolutions on Dataset 1. Starting from a coarse resolution with a size of 25 by 18, the labels sharpen progressively to a fine resolution with a size of 2500 by 1800. This progression aids in

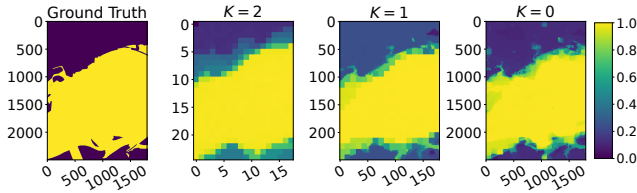


Figure 6: Inferred label at different resolutions (iterations) for dataset 1.

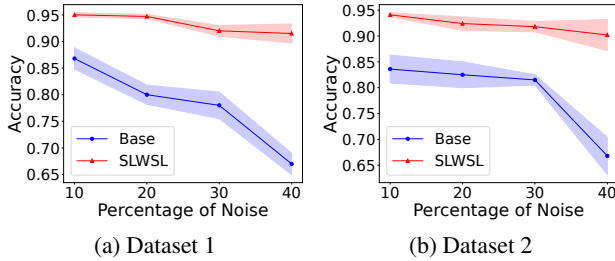


Figure 7: Accuracy comparison under varying noise intensity.

detecting and refining uncertain regions, typically indicative of flood boundaries.

Sensitivity Analysis

To examine the robustness of our proposed SLWSL, especially in the face of noise in observed weak labels, we undertook a sensitivity analysis. For this analysis, we deliberately introduced random noise to the original coarse label, varying the intensity of noise from 10% to 40%. The objective was to evaluate the resultant impact on classification accuracy under progressively challenging conditions. The results of this experiment are detailed in Figure 7. In contrast to the base model, SLWSL remains relatively unaffected by the percentage of noise, maintaining an accuracy rate above 0.9. The underlying reason lies in SLWSL’s unique label inference process: it starts from a coarse resolution and subsequently refines labels, ensuring accurate label inference. The slight variations in the results can be attributed to training randomness and the initialization process.

Ablation Study

To highlight the advantages of the selective refinement strategy, we evaluated the classification performance and computational costs between SLWSL against the variant without selective refinement (w/o SR).

As shown in Table 1, without selective refinement, the accuracy results show that it is slightly less accurate. This can be attributed to too many grounded rules leading to suboptimal convergence in the label inference (E-step) and subsequent neural network training. Table 2 presents a breakdown of the ground atoms, rules, and the time cost for the E-step at the finest resolution, which predominates the training duration. Notably, the E-step’s time requirement is di-

	Dataset 1		Dataset 2	
	SLWSL	w/o SR	SLWSL	w/o SR
Ground Atoms	0.78M	4.50M	8.05M	28.56M
Ground Rules	8.50M	40.47M	72.29M	256.97M
Time Cost (Sec.)	438	3201	7472	44951

Table 2: Comparisons of SLWSL with and without selective refinement at the finest resolution.

rectly related to the number of ground rules, a factor that becomes critical during label inference with a large volume of spatial samples. In contrast, our selective refinement approach grounded only 0.78 million atoms from a possible 4.5 million in Dataset 1, resulting in SLWSL achieving a speed nearly 6.31 times greater than the complete grounding scenario.

Limitations

There are two primary limitations of our SLWSL: 1) Limited scope of spatial knowledge representation: while SLWSL effectively integrates spatial domain knowledge through logic rules, it is limited to pre-defined representations. This constrains its applicability in scenarios where spatial knowledge is conveyed through other formats like geographical knowledge graphs, external spatial statistics/simulation models, or partial differential equations (PDEs). Consequently, the framework may not be effective in contexts where these alternative knowledge representations are essential. 2) Inductive bias and model constraints: the incorporation of logical knowledge as an inductive bias can potentially hinder the learning process. This is due to the constraints it imposes on model training, which might lead to sub-optimal results. Additionally, the reliance on a spatial knowledge database, which may not be comprehensive or entirely accurate for all scenarios, poses a risk of misleading the model.

Conclusion and Future Work

In this work, we presented SLWSL: a spatial-logic-aware weakly supervised learning framework, which integrates deep learning with symbolic spatial logic, enabling precise label inference while training a neural model. By providing a probabilistic framework, informed by observed features, weak labels, and logic constraints, we established a novel method that uniquely addresses the challenges in high-resolution Earth imagery. Central to SLWSL’s efficacy is the utilization of an EM method and a multi-resolution strategy, together to ensure both accuracy and computational efficiency. Our experiments demonstrate that this framework surpasses existing baselines.

While our primary application domain in this work is flood mapping, the proposed method indicates potential utility in other spatial tasks, such as detecting deforestation or monitoring urban expansion. Additionally, incorporating temporal variations from Earth imagery data into our model could enhance its dynamism, making predictions more closely with real-world changes.

Acknowledgments

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. IIS-2147908, IIS-2207072, CNS-1951974, OAC-2152085, the National Oceanic and Atmospheric Administration grant NA19NES4320002 (CISESS) at the University of Maryland. Shigang Chen's work is supported in part by the National Institutes of Health (NIH) grant R01 LM014027.

References

- Bach, S. H.; Broecheler, M.; Huang, B.; and Getoor, L. 2017. Hinge-loss Markov random fields and probabilistic soft logic. *The Journal of Machine Learning Research*, 18(1): 3846–3912.
- Cai, B.; Ding, X.; Chen, B.; Du, L.; and Liu, T. 2022. Mitigating Reporting Bias in Semi-supervised Temporal Commonsense Inference with Probabilistic Soft Logic. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 10454–10462.
- Cao, Y.; and Huang, X. 2022. A coarse-to-fine weakly supervised learning method for green plastic cover segmentation using high-resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188: 157–176.
- Cline, D.; et al. 2009. Integrated water resources science and services: an integrated and adaptive roadmap for operational implementation. *National Oceanic and Atmospheric Administration*.
- Dai, W.-Z.; Xu, Q.; Yu, Y.; and Zhou, Z.-H. 2019. Bridging machine learning and logical reasoning by abductive learning. *Advances in Neural Information Processing Systems*, 32.
- Diligenti, M.; Gori, M.; and Sacca, C. 2017. Semantic-based regularization for learning and inference. *Artificial Intelligence*, 244: 143–165.
- Donadello, I.; Serafini, L.; and Garcez, A. D. 2017. Logic tensor networks for semantic image interpretation. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 1596–1602.
- Duan, X.; Wang, X.; Zhao, P.; Shen, G.; and Zhu, W. 2022. DeepLogic: Joint Learning of Neural Perception and Logical Reasoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4321–4334.
- Eftelioglu, E.; Jiang, Z.; Tang, X.; and Shekhar, S. 2017. The nexus of food, energy, and water resources: Visions and challenges in spatial computing. In *Advances in Geocomputation: Geocomputation 2015–The 13th International Conference*, 5–20. Springer.
- Fabrizio, G.; Farina, A.; and De Maio, A. 2006. Knowledge-based adaptive processing for ship detection in OTH Radar. In *2006 International Radar Symposium*, 1–5. IEEE.
- Foulds, J.; and Frank, E. 2010. A review of multi-instance learning assumptions. *The knowledge engineering review*, 25(1): 1–25.
- Garcez, A. d.; Bader, S.; Bowman, H.; Lamb, L. C.; de Penning, L.; Illumino, B.; Poon, H.; and Zaverucha, C. G. 2022. Neural-symbolic learning and reasoning: a survey and interpretation. *Neuro-Symbolic Artificial Intelligence: The State of the Art*, 342(1).
- Guo, Y.; Wang, F.; Xiang, Y.; and You, H. 2023. Semisupervised Semantic Segmentation With Certainty-Aware Consistency Training for Remote Sensing Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16: 2900–2914.
- He, W.; and Jiang, Z. 2023. A Survey on Uncertainty Quantification Methods for Deep Neural Networks: An Uncertainty Source Perspective. *arXiv preprint arXiv:2302.13425*.
- He, W.; Jiang, Z.; Kriby, M.; Xie, Y.; Jia, X.; Yan, D.; and Zhou, Y. 2022a. Quantifying and Reducing Registration Uncertainty of Spatial Vector Labels on Earth Imagery. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 554–564.
- He, W.; Sainju, A. M.; Jiang, Z.; Yan, D.; and Zhou, Y. 2022b. Earth Imagery Segmentation on Terrain Surface with Limited Training Labels: A Semi-supervised Approach based on Physics-Guided Graph Co-Training. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 13(2): 1–22.
- Hu, Z.; Ma, X.; Liu, Z.; Hovy, E.; and Xing, E. 2016. Harnessing Deep Neural Networks with Logic Rules. In Erk, K.; and Smith, N. A., eds., *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 2410–2420. Berlin, Germany: Association for Computational Linguistics.
- Jiang, Z.; He, W.; Kirby, M. S.; Sainju, A. M.; Wang, S.; Stanislawski, L. V.; Shavers, E. J.; and User, E. L. 2022. Weakly supervised spatial deep learning for earth image segmentation based on imperfect polyline labels. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 13(2): 1–20.
- Jiang, Z.; Zhang, Y.; Adhikari, S.; Yan, D.; Sainju, A. M.; Jia, X.; and Xie, Y. 2023. A Hidden Markov Forest Model for Terrain-Aware Flood Inundation Mapping from Earth Imagery. In *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*, 316–324. SIAM.
- Kimmig, A.; Bach, S.; Broecheler, M.; Huang, B.; and Getoor, L. 2012. A short introduction to probabilistic soft logic. In *Proceedings of the NIPS workshop on probabilistic programming: foundations and applications*, 1–4.
- Li, Q.; Huang, S.; Hong, Y.; Chen, Y.; Wu, Y. N.; and Zhu, S.-C. 2020. Closed loop neural-symbolic learning via integrating neural perception, grammar parsing, and symbolic reasoning. In *International Conference on Machine Learning*, 5884–5894. PMLR.
- Liu, Y.; Li, Q.; Li, X.; He, S.; Liang, F.; Yao, Z.; Jiang, J.; and Wang, W. 2023. Leveraging Physical Rules for Weakly Supervised Cloud Detection in Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*.
- Malkin, K.; Robinson, C.; Hou, L.; Soobitsky, R.; Czawlytko, J.; Samaras, D.; Saltz, J.; Joppa, L.; and Jojic, N. 2018. Label super-resolution networks. In *International Conference on Learning Representations*.

- Manhaeve, R.; Dumancic, S.; Kimmig, A.; Demeester, T.; and De Raedt, L. 2018. Deepproblog: Neural probabilistic logic programming. *Advances in neural information processing systems*, 31.
- Maron, O.; and Lozano-Pérez, T. 1997. A framework for multiple-instance learning. *Advances in neural information processing systems*, 10.
- Mnih, V.; and Hinton, G. E. 2012. Learning to label aerial images from noisy data. In *Proceedings of the 29th International conference on machine learning*, 567–574.
- Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; and Prabhat, f. 2019. Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743): 195–204.
- Richardson, M.; and Domingos, P. 2006. Markov logic networks. *Machine learning*, 62: 107–136.
- Robinson, C.; Hou, L.; Malkin, K.; Soobitsky, R.; Czawlytko, J.; Dilkina, B.; and Jojic, N. 2019. Large scale high-resolution land cover mapping with multi-resolution data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12726–12735.
- Rußwurm, M.; and Korner, M. 2017. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 11–19.
- Tian, J.; Li, Y.; Chen, W.; Xiao, L.; He, H.; and Jin, Y. 2022. Weakly Supervised Neural Symbolic Learning for Cognitive Tasks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 5888–5896.
- Wang, J.; HQ Ding, C.; Chen, S.; He, C.; and Luo, B. 2020a. Semi-supervised remote sensing image semantic segmentation via consistency regularization and average update of pseudo-label. *Remote Sensing*, 12(21): 3603.
- Wang, S.; Chen, W.; Xie, S. M.; Azzari, G.; and Lobell, D. B. 2020b. Weakly supervised deep learning for segmentation of remote sensing imagery. *Remote Sensing*, 12(2): 207.
- Weber, L.; Minervini, P.; Münchmeyer, J.; Leser, U.; and Rocktäschl, T. 2019. NLProlog: Reasoning with Weak Unification for Question Answering in Natural Language. In *57th Annual Meeting of the Association for Computational Linguistics*, 6151–6161. Association for Computational Linguistics.
- Xie, M.; Jiang, Z.; and Sainju, A. M. 2018. Geographical hidden markov tree for flood extent mapping. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2545–2554.
- Xie, Y.; Xu, Z.; Kankanhalli, M. S.; Meel, K. S.; and Soh, H. 2019. Embedding symbolic knowledge into deep networks. *Advances in neural information processing systems*, 32.
- Xu, J.; Zhang, Z.; Friedman, T.; Liang, Y.; and Broeck, G. 2018. A semantic loss function for deep learning with symbolic knowledge. In *International conference on machine learning*, 5502–5511. PMLR.
- Xu, Z.; Xiao, T.; He, W.; Wang, Y.; and Jiang, Z. 2023. Spatial Knowledge-Infused Hierarchical Learning: An Application in Flood Mapping on Earth Imagery. In *The 31st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS)*.
- Yang, W.; Dai, D.; Triggs, B.; and Xia, G.-S. 2012. SAR-based terrain classification using weakly supervised hierarchical Markov aspect models. *IEEE Transactions on Image Processing*, 21(9): 4232–4243.
- Zhou, R.; Zhang, W.; Yuan, Z.; Rong, X.; Liu, W.; Fu, K.; and Sun, X. 2022. Weakly supervised semantic segmentation in aerial imagery via explicit pixel-level constraints. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–17.
- Zhou, Y.; Yan, Y.; Han, R.; Caufield, J. H.; Chang, K.-W.; Sun, Y.; Ping, P.; and Wang, W. 2021. Clinical temporal relation extraction with probabilistic soft logic regularization and global inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 14647–14655.
- Zhou, Z.-H. 2004. Multi-instance learning: A survey. *Department of Computer Science & Technology, Nanjing University, Tech. Rep*, 1.
- Zhou, Z.-H. 2019. Abductive learning: towards bridging machine learning and logical reasoning. *Science China Information Sciences*, 62: 1–3.