ISL: Monitoring Image Segmentation Logic in Medical Imaging Analysis

Ziyan An, Daniel Moyer, Ipek Oguz, Taylor T. Johnson, and Meiyi Ma

Vanderbilt University, Nashville, TN, USA
{ziyan.an, daniel.moyer, ipek.oguz, taylor.johnson, meiyi.ma}@vanderbilt.edu

Abstract. Medical image segmentation plays a critical role in both image interpretation and disease diagnosis. Data-driven approaches such as U-Net have significantly advanced the field by enabling pixel-level classification of anatomical structures. However, the resulting segmentation masks often fail to comply with spatial anatomical constraints. Although properties such as organ connectedness and relative position are well understood by experts, current segmentation practices lack a systematic and explicit means to express, incorporate, or monitor this domain knowledge. In this work, we conduct a systematic study of formal specification rules across diverse medical imaging tasks with domain experts. Based on our analysis, we introduce Image Segmentation Logic (Isl), a novel spatial formalism designed to bridge the gap between expert knowledge and automated analysis. This logic provides an expressive and interpretable framework to specify critical domain knowledge and spatial constraints, and to automatically monitor their satisfaction with segmentation outputs. It integrates variables such as intensity, predicted class labels, confidence scores, and directional relationships, enabling both pixel-level and region-level specification checking. To support practical adoption, we develop an efficient monitoring tool for evaluating segmentation outputs against Isl specifications and formally prove the soundness and correctness of Isl's semantics. We demonstrate the expressiveness and utility of Isl through two case studies on established medical image segmentation datasets. Our results show that Isl enables precise detection of segmentation violations and provides more fine-grained validation than traditional metrics.

Keywords: Specification Language · Medical Image Segmentation · Monitoring Techniques.

1 Introduction

Medical image segmentation plays a vital role in facilitating clinical applications such as disease assessment and diagnosis. Data-driven algorithms, particularly convolutional neural networks such as the U-Net [38] architecture, have significantly advanced segmentation performance and are widely adopted as effective tools in medical imaging [39]. At a micro level, segmentation models generate classification results, also known as masks, for an input image by assigning each pixel to a class such as an organ, lesion, or background. Connected pixels in the segmentation mask form regions, and a well-performing model is expected to comply with spatial rules or constraints defined by biological plausibility, diagnostic accuracy, and model interpretability. As a simple example, for healthy patients, each anatomical structure, such as a lung lobe or an organ, should form a single connected component across the entire image. Additionally, specific positional or directional relationships must hold between segmented regions. For instance, the liver lies below the right lung, and the heart is located between the lungs. Within a region, certain properties must also hold. For example, every boundary pixel of a tumor region should be adjacent to pixels belonging to a specific organ class. Concrete examples of segmentation failures that violate such anatomical and spatial constraints are shown in Fig. 1.

Developing a formal framework to specify and evaluate aforementioned region-level and pixel-level properties and specifications offers several key benefits. First, given the severity and safety-critical nature of data-driven algorithms used for medical imaging segmentation, the European Union's Artificial Intelligence (AI) Act [13] stipulates that high-risk AI systems must ensure robustness, which may involve backup or fail-safe measures. In line with this requirement, a formal framework for medical image segmentation provides an interpretable mechanism to verify and potentially enforce spatial and semantic consistency in model outputs. Second, such a formal framework can serve as an evaluation tool for quantitatively assessing segmentation models beyond standard metrics like IoU, Dice score [11], and Jaccard index [24]. Third, the formal framework can function as a runtime monitor and post-processing tool within clinical pipelines to timely detect violations or anomalies in the outputs of data-driven algorithms. Fourth, for segmentation datasets with existing labels, the formal framework can be used to verify and reassess label quality by identifying potential human errors or inaccuracies caused by automated labeling tools.

However, despite advances in bounding-box-based specification frameworks for image data [12, 22], no existing formal logic framework offers built-in operators that support both pixel-level and region-level reasoning for medical image segmentation, while also integrating learned model attributes (e.g., model confidence, segmentation mask) and directional spatial relations. Therefore, in this paper, we propose a spatial formalism dubbed Image Segmentation Logic (ISL) to specify and evaluate the compliance of medical imaging segmentation algorithms. The design of ISL is motivated by real-world needs in the medical domain, informed by our study of specification requirements across diverse imaging tasks with domain experts, and by the observation that current models lack the means to formally encode or verify such recurring domain knowledge. ISL enables specifications over both pixel-level and region-level properties, defined in terms of image-derived variables such as intensity, anatomical class labels from segmentation, and algorithmic confidence scores, etc. The logic covers key criteria for assessing the usability of medical image segmentation, including specifications about the resulting segmented image, such as the connectivity of a single anatom-

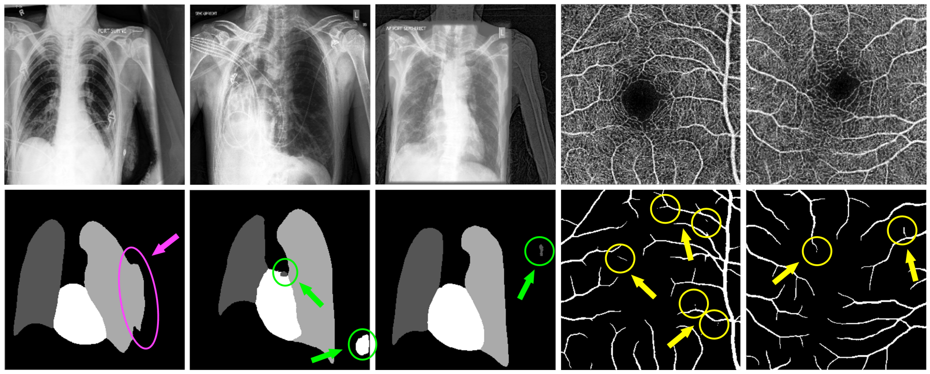

Fig. 1: The first row shows the input image; the second row shows the corresponding predicted segmentation. Violations include incorrect anatomical structure (col. 1), lack of organ coherence (col. 2 and col. 3), and disconnected regions (col. 4 and col. 5). Circled areas and arrows highlight the specific locations where violations occur. Since these constraints are **not** explicitly enforced by standard data-driven models, ISL provides a formal mechanism to monitor such violations.

ical region (e.g., the segmented cochlea must be a single connected component), directional distance between multiple regions (e.g., the left lung is to the left of the heart), as well as specifications about the algorithm itself, such as requiring the model to be at least 95% certain about its output.

To facilitate practical use, we develop both qualitative (Boolean) and quantitative semantics for ISL, and implement an efficient monitoring tool that supports checking segmentation outputs under both semantics. We also provide formal proofs of the logic's soundness and correctness, and derive the computational complexity of the monitoring algorithm. We demonstrate the practicality and comprehensiveness of ISL through two case studies using well-established medical imaging datasets: OCTA-500 [28], which contains retinal images with micron-level resolution, and CheXpert [23], a large-scale, annotated, multi-class segmentation dataset of chest X-rays. The code implementation of ISL is available at https://github.com/AICPS-Lab/isl.git.

Contribution. The key contributions of this paper are summarized as follows. First, we provide an overview of the need for spatial reasoning in medical image segmentation and identify challenges in developing formal specification frameworks. Second, we introduce ISL, a new specification language designed to express and evaluate both pixel-level and region-level properties in image segmentation outputs. The language supports capabilities including connectivity constraints, confidence-based conditions, and relative spatial positioning. Third, we define both qualitative and quantitative semantics for ISL, and develop a monitoring algorithm that implements these semantics to evaluate segmentation outputs. Fourth, we empirically evaluate the ISL monitor on two medical imaging datasets to demonstrate its effectiveness in detecting specification violations and supporting downstream analysis.

2 Segmentation Tasks and Specification Challenges

In this section, we provide background on medical image segmentation tasks, define key data structures associated with them, including the input image and output segmentation mask, and present a comprehensive list of relevant segmentation specifications collected from related literature.

2.1 Medical Image Segmentation Tasks

The use case we consider focuses on the segmentation of anatomical regions such as the brain, left and right lungs, and the heart. Given an input image $\mathbf{I}$, the goal of the segmentation model is to generate a corresponding segmentation mask $\mathcal{S}$ that accurately identifies the targeted semantic label for each pixel.

We consider two-dimensional input images. Each image provided to the segmentation model is represented as $\mathbf{I} \in \mathbb{R}^{H \times W}$, where H and W denote the height and width of the image, respectively. We use f_θ to denote a parameterized segmentation model with learnable parameters θ. Given an input image $\mathbf{I}$, the model outputs a segmentation mask $\hat{y} = f_\theta(\mathbf{I})$, where $\hat{y} \in \mathcal{S}^{H \times W}$. The predicted class label for each pixel $p_{row,col}$ is given by $\mathtt{class}_{row,col} = \arg\max f_\theta(\mathbf{I})_{row,col}$.

The segmentation model is trained using supervised learning methods, where ground-truth annotations $y \in \mathcal{S}^{H \times W}$ are provided either by domain experts or automated labeling tools. The learning objective is to minimize a pixel-wise loss function $\mathcal{L}(\hat{y}, y)$. Two common choices of the loss function are cross-entropy and Dice loss [11]. To optimize the model parameters θ, we adopt a gradient-based optimization algorithm, specifically Adam [26], which iteratively update the parameters until convergence is achieved [45].

2.2 Pixels of Segmentated Images

In our formal logic framework, each image is modeled as a structured collection of pixels, where each pixel is augmented with information derived from both the original input image and the corresponding segmentation mask. We refer to this augmented image as $\mathcal{X}$. Let p denote a pixel in $\mathcal{X}$, and let $p_{row,col}$ represent the specific pixel located at the row and column numbers. We define each pixel in $\mathcal{X}$ to follow the structured format: $p_{row,col} = (\mathtt{row}, \mathtt{col}, \mathcal{I}, \mathtt{class}, \mathtt{prob}, \mathtt{id})$.

Here, $\mathtt{class} \in \{\mathtt{Lung}, \mathtt{Heart}, \mathtt{Background}, \ldots\}$ denotes the predicted semantic label, $\mathtt{id} \in \mathbb{N}$ represents the region identifier to which the pixel belongs, and $\mathtt{prob} \in [0, 1]$ is the associated probability score for that prediction. In this work, we use the maximum softmax probability [16] as a proxy for model confidence in the predicted class. Alternative approaches for estimating or interpreting pixel-level confidence include entropy-based uncertainty [25] and Bayesian sampling methods such as Monte Carlo dropout [35]. For grayscale images considered in this paper, $\mathcal{I}$ refers to the intensity value of each pixel. In the case of RGB images, however, $\mathcal{I}$ can be replaced by a three-dimensional color vector.

In addition to local attributes, each pixel is aware of its four immediate neighbors in the cardinal directions left $(row, col{-}1)$, right $(row, col{+}1)$, up $(row{-}1, col)$, and down $(row{+}1, col)$. Therefore, let $p_{row,col}$ and $p'_{row',col'}$, we have $(p, p') \in \mathsf{Adj}_B$ iff p' is the immediate neighbor of p in direction B, that is:

- $B \in \{\uparrow, \downarrow\}$: $col' = col$, $row' = row{-}1$ if B is $\uparrow$; $row' = row{+}1$ if B is $\downarrow$
- $B \in \{\leftarrow, \rightarrow\}$: $row' = row$, $col' = col{-}1$ if B is $\leftarrow$; $col' = col{+}1$ if B is $\rightarrow$

2.3 Regions of Segmentated Images

Since the image $\mathcal{X}$ includes predicted semantic labels for each pixel, we partition the image into distinct regions, denoted by $\mathcal{R} = \{\tau_1, \tau_2, \ldots, \tau_n\}$, where each region $\tau_i \subseteq \mathcal{X}$ consists of a connected set of pixels sharing the same predicted class label. In practice, each region corresponds to an anatomical or pathological structure, as defined by its semantic class. To distinguish between spatially disjoint regions of the same class, each region is assigned a unique identifier.

We implement a region identifier assignment function to assign unique region IDs to connected regions. Starting from an unvisited pixel, the function performs connected component labeling to grow a region by assigning the same ID to all connected pixels sharing the same class label. This process is repeated across the entire image grid until all class-labeled pixels have been assigned region IDs. As a result, multiple spatially disconnected regions with the same class label are assigned distinct region identifiers.

For any region τ, we define its extremal points as the leftmost and rightmost columns, and the topmost and bottommost rows, that contain pixels belonging to τ. Based on this representation, we define positional relations in the four cardinal directions, above, below, left of, and right of, to characterize spatial relationships within the image space.

2.4 Analysis of Medical Image Segmentation Requirements

To motivate our specification language and identify the requirements of medical image segmentation, we compile and detail specification rules across segmentation tasks involving various anatomical regions, such as the chest and brain.

In Table 1, we list a few key examples of segmentation specifications across different anatomical regions. As shown, the specifications can be categorized into different types. First, for single-class segmentations, properties can be enforced within the target region itself. For example, one can define constraints on the intensity distribution of pixels, such as specifying minimum or maximum allowable intensity values. Second, specifications of the type spatial distance express that a region with certain properties is expected to exist within a specific distance and direction relative to a target region. These constraints define the spatial relationship between two anatomical regions within the segmentation mask, describing both directional position and proximity. Such specifications are common in

Table 1: Example segmentation specifications for various anatomical regions.

Region	Type	Textual Description	Ref.
Chest	Spatial distance	For segmented heart region, there exists a left lung to the left and a right lung to the right within a fixed distance.	[9]
Lung	Spatial distance	For segmented left lung region, there exists a right lung region to the right that does not overlap with it.	[17]
Lung	Region class label	For all pixels segmented as trachea, they do not belong to either left or right lung regions.	[40]
Brain	Spatial symmatry	For segmented left hemisphere region, there exists a corresponding right hemisphere to the right such that the class labels are the same.	[49]
Retinal	Region class label	For segmented FAZ region, all pixels within it are not labeled as vessel, and for segmented artery region, no pixel within it is labeled as vein.	[20]
Cochlea	Spatial distance	For segmented cochlea region, there exists a vestibule region posterior to it within some pixels.	[36]
Cochlea	Region intensity	For segmented cochlea region, at least a certain percentage of pixels have intensity between a specified range.	[30]
Tumor	Region intensity	For segmented tumor region, no pixel within it has intensity within the typical cerebrospinal fluid (CSF) range.	[34]

anatomical areas like the chest, lungs, and eyes, where the relative arrangement of structures is clinically well-defined. The type spatial symmetry, also describing the relation between two regions, can be viewed as a special case of spatial distance, where there exists a corresponding region at a fixed mirrored distance along a defined axis. These observations inform the design of Isl, which lead to a specification language that encompasses the diverse constraints observed in medical segmentation tasks. This includes intra-region properties such as "region intensity" and inter-region relationships such as "spatial distance" and "spatial symmetry." The examples also highlight the need for a unified formalism capable of expressing a broad range of clinically meaningful properties, both within individual regions and between multiple anatomical structures.

3 Image Segmentation Logic (Isl)

Targeting the above-mentioned unique challenge, we propose a new specification language dubbed Image Segmentation Logic (Isl) to reason about medical image segmentation tasks. Figure 2 illustrates the overall Isl framework. In this section, we first present the syntax and the semantics of the Isl specification language.

Building on the data structures defined for pixels and regions in the augmented image $\mathcal{X}$, we now formally define the spatial model over which the Isl logic is interpreted. The image model is denoted as $\mathcal{M} = ((\mathcal{X}, \mathcal{R}), \mathcal{V})$, where $\mathcal{V}$ is a valuation function that maps each pixel p to a tuple of attributes: $\mathcal{V}(p) = (\mathcal{I}(p), \mathtt{class}(p), \mathtt{prob}(p), \mathtt{id}(p), \mathtt{row}(p), \mathtt{col}(p))$. In addition, we define pixel objects and region objects as two types of input objects of Isl. Specifically, a pixel object is defined by its coordinate position $p \coloneqq (\mathtt{row}, \mathtt{col}) \in \mathcal{X}$. A region object $\tau \coloneqq (p_{\mathrm{tl}}, p_{\mathrm{tr}}, p_{\mathrm{bl}}, p_{\mathrm{br}}) \in \mathcal{R}$ is represented by the coordinates of its four corners: top-left, top-right, bottom-left, and bottom-right.

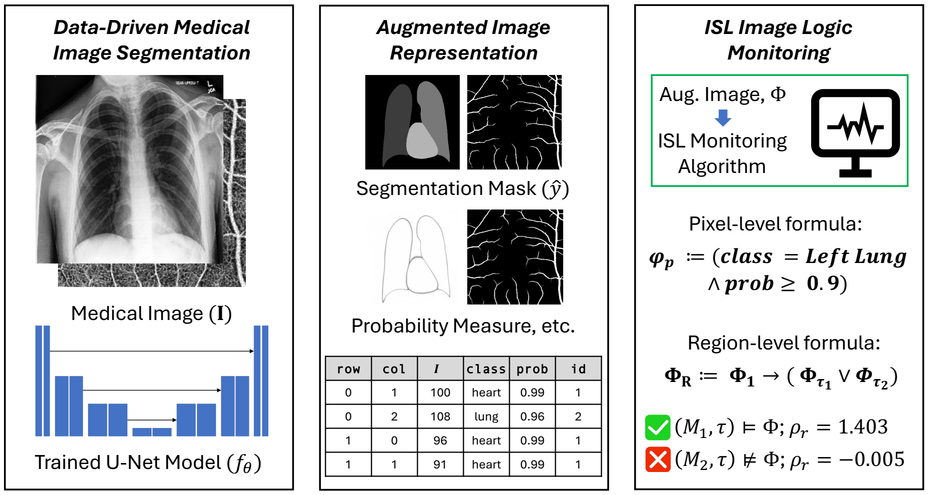

Fig. 2: Overview of the IsL framework for specification monitoring in medical image segmentation. Framework begins with images segmented into class-labeled regions by a model. Each output is augmented with pixel-level attributes. IsL then evaluates logic specifications encoding anatomical priors and spatial rules, producing Boolean and robustness scores to detect structural violations.

3.1 Syntax

To support both intra-region and inter-region reasoning over medical image segmentations, we define the syntax of IsL in a hierarchical manner, with separate formulas for pixel-level and region-level specifications. Pixel-level formulas enable reasoning about local attributes such as intensity and class labels, while region-level formulas support spatial reasoning across different segmented regions. These two levels are connected using quantifiers and spatial operators, which aggregate pixel-level predicates into region-level statements.

Definition 1 (Syntax of IsL). *The pixel-level and region-level formulas of IsL are defined as:*

$$\varphi ::= \top \mid \pi \sim d \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \bigcirc_B \varphi$$
$$\Phi ::= \boxed{\exists}\,\varphi \mid \boxed{\forall}\,\varphi \mid \neg\Phi \mid \Phi \wedge \Phi \mid \Phi \vee \Phi \mid \blacktriangleright_s \Phi \mid \triangleright_s \Phi$$

In this definition, φ denotes a pixel-level formula, and Φ denotes a region-level formula. The constant $\top$ is the Boolean value *True*, $\sim \in \{=, \geq, \leq\}$, $s \in \mathbb{R}_{\geq 0}$, and $B \in \{\uparrow, \downarrow, \leftarrow, \rightarrow\}, B \neq \emptyset$. The operators $\neg, \wedge, \vee$ follow the standard first-order semantics. The implication operator $\varphi \rightarrow \varphi$ is equivalent to $\neg\varphi \vee \varphi$.

Atomic predicates of a pixel-level formula φ take the form $\pi \sim d$, where d is a numeric or symbolic constant, and π is a pixel-level attribute defined as:

$$\pi ::= \texttt{id} \mid \texttt{class} \mid \texttt{prob} \mid \mathcal{I} \mid \texttt{row} \mid \texttt{col}$$

The predicate $\texttt{id} = d$ specifies that pixel p is assigned the region identifier d, while $\texttt{class} = d$ indicates that p is labeled with the semantic class $d \in \mathcal{C}$. The predicate $\texttt{prob} \sim d$ compares the predicted confidence score of p to a threshold d using a relational operator. The term $\mathcal{I} \sim d$ constrains the pixel's intensity value, while $\texttt{row} \sim d$ and $\texttt{col} \sim d$ impose spatial constraints on the pixel's vertical and horizontal coordinates, respectively. In addition, the operator $\bigcirc_B \varphi$ is a directional pixel-level operator that evaluates whether the formula φ holds at the neighboring pixel of a given pixel p in the directions specified by B. Intuitively, $\bigcirc_B$ moves to the next pixel from the current pixel in direction B and checks whether φ is satisfied there.

Region-level formulas Φ allow reasoning over segmented regions through quantification and directional operators. The boxed quantifiers $\boxed{\forall}\varphi$ and $\boxed{\exists}\varphi$ enable the assertion of the existence or universality of the pixel-level property φ. Region formulas also support logical negation, conjunction, and disjunction with $\neg$, $\wedge$, and $\vee$ operators, respectively. Directional spatial relations between regions are expressed using a family of direction-specific operators, each annotated with a subscript s that specifies the distance threshold. The operator $\blacktriangleright_s \Phi$ evaluates whether there exists a region located strictly to the right, at a directional Manhattan distance of at least s, in which the spatial property Φ holds. Similarly, $\blacktriangleleft_s$, $\blacktriangle_s$, and $\blacktriangledown_s$ correspond to left, up, and down each directional relations, respectively. Two variants of each directional operator are supported. Specifically, ISL includes a strict version (i.e., $\blacktriangleright_s$), which requires the entirety of the target region to lie in the specified direction, and a weak version (i.e., $\triangleright_s$), which allows partial overlap as long as some part of the region lies in the given direction.

Running Example. We illustrate how ISL can express local connectivity and directional constraints. Consider the specification: "a segmented class (e.g., left lung) must appear as a single connected region." This requires that the segmentation does not contain multiple high-confidence, spatially disconnected components with the same class label. We define a pixel-level formula as $\varphi_1 := (\texttt{class} = \text{LeftLung}) \wedge (\texttt{prob} \geq 0.9) \wedge (\texttt{id} = \texttt{id1})$, asserting that a pixel belongs to a confident left lung region $\texttt{id1}$. The region-level formula $\Phi_1 := \boxed{\forall}\varphi_1$ ensures all pixels in region $\texttt{id1}$ satisfy these conditions. To enforce uniqueness, we require that no other region satisfies the same predicate. For each region $\tau \neq \texttt{id1}$, let $\varphi_\tau := (\texttt{class} = \text{LeftLung}) \wedge (\texttt{prob} \geq 0.9) \wedge (\texttt{id} = \tau)$ and $\Phi_\tau := \boxed{\forall}\varphi_\tau$. The uniqueness constraint is then $\Phi_{\text{uniqueness}} := \neg\,(\Phi_1 \wedge (\vee_{\tau \neq \texttt{id1}} \Phi_\tau))$.

As a second example, we can specify a directional constraint between two anatomical structures as follows: $\Phi_{\text{direction}} := \neg(\Phi_{\text{LeftLung}}) \vee \blacktriangleright_{20} \Phi_{\text{RightLung}}$, where $\varphi_{\text{LeftLung}} := (\texttt{class} = \text{LeftLung})$, $\varphi_{\text{RightLung}} := (\texttt{class} = \text{RightLung})$, and $\Phi_{\text{LeftLung}} := \boxed{\forall}\varphi_{\text{LeftLung}}$, $\Phi_{\text{RightLung}} := \boxed{\forall}\varphi_{\text{RightLung}}$. This formula asserts that if a region exists in which every pixel is labeled as `LeftLung`, then there must also exist a region labeled as `RightLung` that lies strictly to its right, with a minimum directional Manhattan distance of 20 pixels between them.

3.2 Qualitative Semantics

In the following sections, we present the qualitative and quantitative semantics of Isl. The qualitative semantics returns a Boolean truth value, while the quantitative semantics returns a real number that measures the degree of satisfaction for a given image model. In Definition 2, we first present the Boolean, or qualitative, semantics of Isl. Given an image model $\mathcal{M}$ and either a pixel object p or a region object τ, the truth relation is defined as follows:

Definition 2 (Qualitative Semantics of Isl).

$$
\begin{aligned}
(\mathcal{M},\tau) &\models \boxed{\exists}\,\varphi \Leftrightarrow \exists p \in \tau :\ (\mathcal{M},\tau,p) \models \varphi\\
(\mathcal{M},\tau) &\models \boxed{\forall}\,\varphi \Leftrightarrow \forall p \in \tau :\ (\mathcal{M},\tau,p) \models \varphi\\
(\mathcal{M},\tau) &\models \neg\Phi \Leftrightarrow (\mathcal{M},\tau) \not\models \Phi\\
(\mathcal{M},\tau) &\models \Phi_1 \wedge \Phi_2 \Leftrightarrow (\mathcal{M},\tau) \models \Phi_1 \ \textit{and}\ (\mathcal{M},\tau) \models \Phi_2\\
(\mathcal{M},\tau) &\models \Phi_1 \vee \Phi_2 \Leftrightarrow (\mathcal{M},\tau) \models \Phi_1 \ \textit{or}\ (\mathcal{M},\tau) \models \Phi_2\\
(\mathcal{M},\tau) &\models \blacktriangleright_s \Phi \Leftrightarrow \exists \tau' \in \mathsf{AlignedRegions}(\tau,s) :\ (\mathcal{M},\tau') \models \Phi\\
(\mathcal{M},\tau) &\models \triangleright_s \Phi \Leftrightarrow \exists \tau' \in \mathsf{AlignedRegionsWeak}(\tau,s) :\ (\mathcal{M},\tau') \models \Phi\\
(\mathcal{M},\tau,p) &\models \top \Leftrightarrow \textit{True}\\
(\mathcal{M},\tau,p) &\models \pi \sim d \Leftrightarrow \delta_{bool}(\pi,\sim,d,p)\\
(\mathcal{M},\tau,p) &\models \neg\varphi \Leftrightarrow (\mathcal{M},\tau,p) \not\models \varphi\\
(\mathcal{M},\tau,p) &\models \varphi_1 \wedge \varphi_2 \Leftrightarrow (\mathcal{M},\tau,p) \models \varphi_1 \ \textit{and}\ (\mathcal{M},\tau,p) \models \varphi_2\\
(\mathcal{M},\tau,p) &\models \varphi_1 \vee \varphi_2 \Leftrightarrow (\mathcal{M},\tau,p) \models \varphi_1 \ \textit{or}\ (\mathcal{M},\tau,p) \models \varphi_2\\
(\mathcal{M},\tau,p) &\models \bigcirc_B\, \varphi \Leftrightarrow \exists p' : (p,p') \in \mathsf{Adj}_B \ \textit{and}\ (\mathcal{M},\tau,p') \models \varphi
\end{aligned}
$$

In this definition, the satisfaction of atomic predicates $\pi \sim d$ at pixel p is determined by a Boolean evaluation function δ_{bool}, defined as follows:

$$
\delta_{\text{bool}}(\pi,\sim,d,p) = \begin{cases} [\pi](p) = d & \text{if } \pi \in \{\texttt{id},\texttt{class}\}\\ [\pi](p) \sim d & \text{if } \pi \in \{\texttt{prob},\mathcal{I},\texttt{row},\texttt{col}\} \end{cases}
$$

To define directional spatial relationships between regions, we introduce the concept of aligned regions. Given a prototype region specified by four labeled corner points (top-left, top-right, bottom-left, bottom-right), we construct a sequence of translated regions by shifting this prototype away from a reference region in a specified cardinal direction by a minimum step size s.

We denote the set of aligned regions at distance s as $\mathsf{AlignedRegions}(\tau,s)$, where each candidate region is obtained by incrementally translating the prototype region's corner points in the chosen direction. The translation continues as long as the resulting region remains within the image bounds. For each translated region, we check whether there exists a segmented component in the image whose corner points are fully enclosed within the translated rectangle. These enclosed components are collected as the valid aligned regions. The directional satisfaction operator $\blacktriangleright_s \Phi$ then evaluates whether the region-level formula Φ holds on at least one such aligned region. In addition, to support more flexible specifications, we define a relaxed variant $\mathsf{AlignedRegionsWeak}(\tau,s)$, which includes all segmented components that intersect the translated prototype region.

3.3 Quantitative Semantics

In Definition 3, we present the quantitative semantics of Isl, where ρ_p denotes the satisfaction degree for pixel-level formulas and ρ_r denotes the satisfaction degree for region-level formulas. Intuitively, the quantitative semantics reflect how strongly the image components satisfy or violate a given formula.

Definition 3 (Quantitative Semantics of Isl).

$$\begin{aligned}
\rho_r(\boxed{\exists}\,\varphi,\tau) &= \max\nolimits_{p\in\tau}\rho_p(\varphi,p)\\
\rho_r(\boxed{\forall}\,\varphi,\tau) &= \min\nolimits_{p\in\tau}\rho_p(\varphi,p)\\
\rho_r(\neg\Phi,\tau) &= -\rho_r(\Phi,\tau)\\
\rho_r(\Phi_1\wedge\Phi_2,\tau) &= \min(\rho_r(\Phi_1,\tau),\rho_r(\Phi_2,\tau))\\
\rho_r(\Phi_1\vee\Phi_2,\tau) &= \max(\rho_r(\Phi_1,\tau),\rho_r(\Phi_2,\tau))\\
\rho_r(\blacktriangleright_s\Phi,\tau) &= \max\nolimits_{\tau'\in\mathsf{AlignedRegions}(\tau,s)}\rho_r(\Phi,\tau')\\
\rho_r(\triangleright_s\Phi,\tau) &= \max\nolimits_{\tau'\in\mathsf{AlignedRegionsWeak}(\tau,s)}\rho_r(\Phi,\tau')\\
\rho_p(\top,p) &= +\infty\\
\rho_p(\pi\sim d,p) &= \delta_{quant}(\pi,\sim,d,p)\\
\rho_p(\neg\varphi,p) &= -\rho_p(\varphi,p)\\
\rho_p(\varphi_1\wedge\varphi_2,p) &= \min(\rho_p(\varphi_1,p),\rho_p(\varphi_2,p))\\
\rho_p(\varphi_1\vee\varphi_2,p) &= \max(\rho_p(\varphi_1,p),\rho_p(\varphi_2,p))\\
\rho_p(\bigcirc_B\varphi,p) &= \begin{cases}\rho_p(\varphi,p') & \textit{if } (p,p')\in\mathsf{Adj}_B \textit{ and } p'\in\mathcal{X}\\ -\infty & \textit{otherwise}\end{cases}
\end{aligned}$$

In the quantitative semantics, atomic predicates are evaluated using a scoring function δ_{quant}, which measures the degree to which a pixel satisfies the specified condition and is defined as follows:

$$\delta_{\text{quant}}(\pi,\sim,d,p)=\begin{cases}1 & \text{if } \pi\in\{\texttt{id},\texttt{class}\} \text{ and } [\pi](p)=d\\ -1 & \text{if } \pi\in\{\texttt{id},\texttt{class}\} \text{ and } [\pi](p)\neq d\\ [\pi](p)-d & \text{if } \sim \text{ is } \geq \text{ and } \pi\in\{\texttt{prob},\mathcal{I},\texttt{row},\texttt{col}\}\\ d-[\pi](p) & \text{if } \sim \text{ is } \leq \text{ and } \pi\in\{\texttt{prob},\mathcal{I},\texttt{row},\texttt{col}\}\end{cases}$$

Running Example. Using the examples above, we illustrate how Isl semantics are applied. For Φ_1, the Boolean semantics evaluate to true if every pixel in region `id1` satisfies the class, confidence, and region identifier conditions. The quantitative semantics compute a robustness score by measuring values such as the margin between each pixel's confidence and the threshold of 0.9, with the overall region score given by the minimum margin across all pixels. For the uniqueness constraint $\Phi_{\text{uniqueness}}$, Boolean satisfaction holds if no other region besides `id1` satisfies the same conjunction. Quantitatively, violation can be measured by the maximum robustness score among all conflicting regions $\tau\neq\texttt{id1}$. For the directional constraint $\Phi_{\text{direction}}$, the semantics check whether there exists a region satisfying $\Phi_{\text{RightLung}}$ that lies at least 20 pixels to the right of a region satisfying Φ_{LeftLung}. We first construct a set of aligned regions translated in the specified direction by the required distance. For each aligned region, we evaluate whether it satisfies $\Phi_{\text{RightLung}}$, and if so, compute its robustness score. The final quantitative score is the maximum robustness value among all candidate regions.

3.4 Semantic Properties of Isl

We define two key properties of the quantitative semantics of Isl, namely soundness and correctness. Intuitively, the soundness property states that the sign of the quantitative semantics corresponds to the Boolean truth value. Specifically, a positive quantitative measure ρ indicates satisfaction, while a negative value indicates violation. The correctness property states that if two images differ by a bounded distance, and one of them satisfies a given property, then the other image also satisfies the same property, if the distance between them is less than the robustness value of the satisfying image.

Theorem 1 (Soundness). *We define the soundness property of Isl as follows. Let Φ be an Isl region-level formula and τ be a region in image model $\mathcal{M}$. Then:*

$$(\mathcal{M}, \tau) \models \Phi \Rightarrow \rho_r(\Phi, \tau) > 0 \quad \textit{and} \quad (\mathcal{M}, \tau) \not\models \Phi \Rightarrow \rho_r(\Phi, \tau) < 0.$$

Let $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ be two image models defined over the same spatial domain $\mathcal{X}$: $\mathcal{M}^{(1)} = ((\mathcal{X}, \mathcal{R}), \mathcal{V}^{(1)}), \quad \mathcal{M}^{(2)} = ((\mathcal{X}, \mathcal{R}), \mathcal{V}^{(2)})$. Let $\tau \in \mathcal{R}$ be a region object common to both models. Define the per-pixel attribute distance as $\delta(\mathcal{V}^{(1)}(p), \mathcal{V}^{(2)}(p)) = \max_i d_i \left(v_i^{(1)}(p), v_i^{(2)}(p)\right)$, where $v_i(p)$ denotes the i-th attribute in $\mathcal{V}(p)$, and d_i is either an absolute difference or an indicator function. Let the model-wise distance be: $d_\infty(\mathcal{V}^{(1)}, \mathcal{V}^{(2)}) := \sup_{p \in \mathcal{X}} \delta(\mathcal{V}^{(1)}(p), \mathcal{V}^{(2)}(p))$.

Theorem 2 (Correctness). *We define the correctness property of Isl region-level formulas as follows. If the following conditions hold:*

$$(\mathcal{M}^{(1)}, \tau) \models \Phi, \rho_r(\Phi, \tau; \mathcal{M}^{(1)}) > 0, d_\infty(\mathcal{V}^{(1)}, \mathcal{V}^{(2)}) < \rho_r(\Phi, \tau; \mathcal{M}^{(1)})$$

Then, $(\mathcal{M}^{(2)}, \tau) \models \Phi$.

4 Monitoring Algorithm for Isl

4.1 Monitoring of Segmented Images

Algorithm 1 presents the pseudocode for the quantitative monitoring algorithm of Isl, implemented based on the logic's recursive quantitative semantics. A corresponding qualitative monitoring algorithm, which returns a Boolean satisfaction value based on the qualitative semantics, is also developed but omitted here for brevity. Below, we describe the quantitative monitoring algorithm and its key components. The monitoring algorithm begins by parsing the logical formula and recursively computing the quantitative scores of its subformulas, aggregating the results using the corresponding semantic operators. For spatial operators, the algorithm updates the evaluation context by traversing aligned or adjacent structures within the image.

In Algorithm 1, the `EvalPixelAtoms` function evaluates pixel-level atomic predicates at a given pixel location by querying the valuation function according to the predicate type specified in the formula. The `NeighborByDirection`

Algorithm 1 Isl Quantitative Scoring Function

```
Require: Φ: Isl formula, object: pixel or region, image: augmented image
function Score(Φ, object, image)
    if Φ is PixelAtomic then
        return EvalPixelAtoms(Φ, object, image)
    else if Φ is Next then
        p ← object
        p′ ← NeighborByDirection(p, Φ.direction)
        return Score(Φ, p′, image)
    else if Φ is Exists or Forall then
        τ ← object
        pixels ← EnumeratePixelsInRegion(τ)
        if Φ is Exists then
            return sup_{p∈pixels} Score(Φ, p, image)
        else if Φ is Forall then
            return inf_{p∈pixels} Score(Φ, p, image)
        end if
    else if Φ is StrongDistance or WeakDistance then
        τ ← object
        if Φ is StrongDistance then
            regions ← AlignedRegions(τ, Φ.direction, Φ.s)
        else
            regions ← AlignedRegionsWeak(τ, Φ.direction, Φ.s)
        end if
        return sup_{τ′∈regions} Score(Φ, τ′, image)
    else if Φ is Negation then
        return −Score(¬Φ, object, image)
    else if Φ is And then
        return min(Score(Φ1, object, image), Score(Φ2, object, image))
    else if Φ is Or then
        return max(Score(Φ1, object, image), Score(Φ2, object, image))
    end if
end function
```

function returns the adjacent pixel of the current pixel in the direction specified by B. The `EnumeratePixelsInRegion` function enumerates all pixels enclosed within the queried region τ. Finally, the `AlignedRegions` function returns all region candidates that are spatially aligned with the queried region τ in the direction and at the distance specified by the operator and parameter s.

4.2 Complexity Analysis

The overall time complexity of the Isl monitoring algorithm depends on the image size, the structure of the logic formula, and the number of segmented regions. We define $R = |\mathcal{R}|$ as the total number of regions identified by the segmentation algorithm, m as the maximum number of pixels in any region, s as the maximum number of aligned regions reachable from a given region, and S_φ as the total number of nodes in the syntax tree of the monitored formula.

By the semantics definition, the time complexity for evaluating operators $\pi \sim d$, $\neg$, $\wedge$, and $\vee$ is $O(1)$, as they only require constant-time operations. The pixel-level operator $\bigcirc_B$ also incurs constant time $O(1)$, as it accesses a neighboring pixel based on a fixed direction. For the quantifiers $\boxed{\exists}$ and $\boxed{\forall}$, the worst-case complexity is $O(m)$. For spatial reachability operators $\blacktriangleright_s$, the worst-case complexity is $O(s)$. Therefore, for one Isl formula monitoring a single object in the image, the time complexity is bounded by $O(S_\varphi \times \max\{m, s\})$. For an Isl formula monitoring every region across the entire image, the overall worst-case time complexity is $O(R \times S_\varphi \times \max\{m, s\})$.

5 Case Studies

To evaluate the usefulness and efficiency of Isl and its monitoring algorithm, we conduct two case studies using state-of-the-art medical image segmentation datasets. The objectives of the case studies are twofold: (1) to demonstrate whether monitoring image segmentation properties in multi-class segmentations provides additional safety-relevant information beyond traditional metrics for state-of-the-art segmentation models; (2) to investigate whether the properties specified by Isl accurately capture the spatial requirements of the segmentation mask, and whether the monitor can faithfully identify violations.

To assess the overhead of our monitoring approach, we conduct a runtime evaluation of the Isl monitoring algorithm across a range of formula types. We implemented both the quantitative and qualitative versions of Algorithm 1 for use in our case studies. We tested the monitoring algorithm on a MacOS Sequoia system equipped with an Apple M2 chip. All deep learning models for medical image segmentation were trained and evaluated on a Linux PC running Ubuntu 18.04, equipped with an Intel Core i9-10850K CPU and an NVIDIA GeForce RTX 3070 GPU. Results show that even the most complex multi-region quantitative specifications average under 280 milliseconds per image, while simpler pixel-level formulas execute in under 10 microseconds.

5.1 Case 1: Single Class Segmentation Connectivity

In the first case study, we examine the OCTA-500 dataset [27] and evaluate the connectivity of large retinal vessels, which is considered one of the most difficult spatial requirements in medical image segmentation [6, 41]. We evaluate four U-Net models. Two of them are trained using pixel-wise binary cross-entropy loss, and the other two are trained with an additional Betti matching loss that encourages topological consistency between predictions and ground truth.

To formalize vessel connectivity, we define an Isl specification that enforces border contact for each predicted vessel region. Anatomically, large vessels enter the retina through the image boundary. There should be no standalone vessel segments entirely within the center of the image. We define the pixel-level formula as $\varphi_r = (\texttt{class} = \texttt{vessel})$ and the corresponding region-level formula as $\Phi_r = \boxed{\forall}\varphi_r$. The border contact requirement is expressed as $\Phi_r^{\text{border}} = \boxed{\exists}(\texttt{row} = 0) \vee \boxed{\exists}(\texttt{col} = 0) \vee \boxed{\exists}(\texttt{row} = 303) \vee \boxed{\exists}(\texttt{col} = 303)$. The complete specification is written as $\Phi_r^{\text{connect}} = \Phi_r \rightarrow \Phi_r^{\text{border}}$, which ensures that any segmented vessel region with high confidence must intersect the image border.

In Table 2, we present the results of evaluating the specification Φ_r^{connect} on the test set of 20 images from the OCTA dataset. Specifically, the violation rate refers to the percentage of regions that violate the border connectivity

Table 2: Performance comparison of segmentation models for OCTA-500.

Seg. Model	Vio. Rate (%)	Vio. Count	Dice↑	mIoU↑
U-Net-m1	33.71	238	0.92	0.85
U-Net-m2	30.63	208	0.92	0.85
U-Net-Betti-m1	24.34	148	0.91	0.84
U-Net-Betti-m2	32.23	223	0.92	0.85
Ground Truth	5.84	29	1.00	1.00

specification among all identified regions in the segmentation mask, while the violation count is the total number of such violations across the entire test set. The Dice score and mIoU represent the mean pixel-level performance metrics computed over the test set. In addition, we assess the structural quality of the provided ground truth annotations using the same connectivity specification.

While all U-Net variants achieve similar pixel-level performance in terms of Dice and mIoU scores, their violation rates differ substantially. Notably, U-Net-Betti-m1 achieves the lowest violation count (148), suggesting some benefit from topological supervision through Betti matching loss. However, the inconsistency across models indicates that Betti loss alone is insufficient to ensure structural correctness. These results also highlight that structural validation metrics, such as region-level violation rate, offer complementary insights beyond traditional pixel-wise metrics. Despite consistently high Dice and mIoU scores, violation counts vary by up to 90 regions across models, demonstrating that standard metrics do not reliably enforce topological fidelity. Interestingly, the ground truth exhibits a violation count of 29, highlighting that structural imperfections exist in manual annotation errors. Overall, this case study indicates that the integration of logic-based constraints using Isl provides more targeted and anatomically meaningful supervision.

5.2 Case 2: Multi-Class Segmentation Spatial Alignment

In the second case study, we evaluate multi-class chest organ segmentation using the CheXmask dataset [14], which provides pixel-level anatomical segmentation masks for chest X-ray images. These masks were generated by the HybridGNet model and refined using the Reverse Classification Accuracy (RCA) technique for quality control [15]. This combination ensures consistent and high-quality segmentations across multiple public datasets, including CheXpert [23]. In our study, we used the CheXmask-provided segmentations as ground truth and trained a U-Net model using a combination of Dice loss and cross-entropy loss. Model performance was evaluated using the Dice Similarity Coefficient (DSC) for both lung and heart structures. The test set contains 159 images in total.

We evaluate spatial logic specifications that capture left–right anatomical relationships. We define an Isl specification that enforces the expected spatial arrangement in which the right lung lies to the right of the left lung. More specifically, if a region is classified as the right lung, there must exist another

Table 3: Evaluation of Isl specifications on CheXmask multi-class dataset.

Isl Formula	**Target Variable**	**Semantics**	**Results**
Region distance $\blacktriangleright_s$	L. lung, R. lung	Quanlitative	99%
Weak distance $\triangleright_s$	R. lung, Heart	Quanlitative	99%
Universality $\boxed{\forall}$	Heart, Max Intensity	Quantitative	1.403
Universality $\boxed{\forall}$	Heart, Min Intensity	Quantitative	9.996

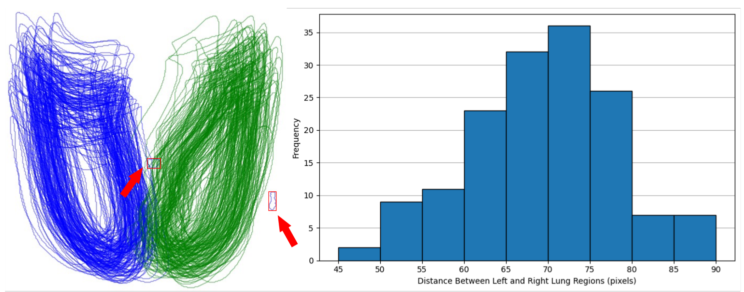

Fig. 3: Evaluation of directional spatial separation between left and right lungs.

region classified as the left lung that lies a sufficient distance to its left to ensure spatial separation. In addition to lung positioning, we also define a spatial relationship between the heart and the right lung using a weak directional distance constraint. Moreover, we include two specifications that assess pixel intensity distributions within the segmented heart. Specifically, we use universal quantification to verify that the maximum and minimum intensity values within the heart fall within expected bounds, where the expected values are derived from the actual intensity distribution of the heart region in the ground truth mask.

Table 3 shows the evaluation outcomes for four Isl specifications. The first two specifications, based on directional distance operators, are satisfied with 99% accuracy. The violated test images are marked by red arrows and boxes in the left subfigure of Figure 3, where the visualization confirms that the Isl monitor correctly identifies the violating cases. The remaining two specifications assess pixel intensity properties within the heart using universal quantification. The average quantitative satisfaction scores for maximum and minimum intensity bounds are 1.403 and 9.996, respectively. These values indicate that the model preserves reasonable intensity consistency, which supports anatomical plausibility, as different organs exhibit distinct radiodensity characteristics.

To assess whether model predictions reflect anatomically plausible spatial separation between the lungs, we defined a region-level Isl specification requiring the right lung to lie to the right of the left lung by at least d pixels. For each image, we computed the minimum value of d for which the specification was satisfied. The results shown in the right subfigure of Figure 3 indicate that most segmentations satisfy the constraint at relatively low distance thresholds, which is consistent with the visualizations in the left subfigure.

Table 4: Average monitoring time per image for different ISL formula types.

Formula Type	Image Size	R Count	Semantics	Avg. Time (s)
Nested Pixel	320×320	N/A	Qualitative	2e-5 ± 7e-6
Nested Pixel	320×320	N/A	Quantitative	2e-5 ± 4e-6
Single Region	320×320	4	Qualitative	0.04 ± 0.00
Single Region	320×320	4	Quantitative	0.07 ± 0.01
Distance	320×320	4	Qualitative	0.18 ± 0.04
Distance	320×320	4	Quantitative	0.24 ± 0.05
Nested Pixel	304×304	N/A	Qualitative	7e-6 ± 0e-6
Nested Pixel	304×304	N/A	Quantitative	8e-6 ± 1e-6
Single Region	304×304	30	Qualitative	0.05 ± 0.00
Single Region	304×304	30	Quantitative	0.05 ± 0.00
Multi-Region	304×304	30	Qualitative	0.15 ± 0.02
Multi-Region	304×304	30	Quantitative	0.28 ± 0.01

5.3 Runtime Evaluation

We assess the computational efficiency of the ISL monitoring framework by measuring the average runtime per image for different formula types and semantics. Table 4 summarizes the results across both chest X-ray images (320×320) and retinal OCTA images (304×304). These results demonstrates that both formula structure and semantic type affect monitoring time. Pixel-level formulas, whether evaluated qualitatively or quantitatively, exhibit negligible runtime due to their simplicity and localized computation. In contrast, region-based specifications incur higher runtime costs due to component extraction and region-wise logical evaluation. Specifically, single-region formulas take 40–70 milliseconds per image, while directional distance constraints, which require spatial reasoning across regions, are the most expensive, averaging up to 240 milliseconds for quantitative semantics. For OCTA data with higher region counts (up to 30 per image), multi-region implications further increase runtime, especially under quantitative evaluation. These results demonstrate that while ISL enables expressive specification of spatial constraints, the associated computational overhead remains acceptable for both online and offline use cases.

6 Related Work

Two major lines of work have shaped formal reasoning for image data. The first focuses on topological properties such as connectedness and concavity [4, 29, 37], while the second addresses geometric relations like position and orientation [7, 19, 22]. RCC-8 [37] and $S4_u$ [4] form the foundation of topological reasoning, with extensions adding Boolean combinations [46] and temporal support [5]. On the geometric side, DAC [7] and SpaTeL [19] enable reasoning over directional or rectangular regions, while logics such as STPL [22] and STSL [29] adapt to grid-based visual domains. SaSTL [31] extends these frameworks with spatial aggregation and counting, and builds on STL [33]. These logics have been applied to structured reasoning over pixel-wise and region-wise properties in image segmentation and layout validation [18]. More recently, formal specification frameworks that leverage scene graphs to monitor semantic-level properties have emerged to bridge the gap between visual data and logical reasoning [42, 47, 48].

Formal logic has also been applied to monitor and guide learning-enabled systems in recent literature [2, 10]. STLnet [32] and FedSTL [1] monitor RNN and federated models via STL-based runtime checks. TQTL [12] targets safety-critical perception in autonomous driving, while An et al. [3] apply logic to multi-level driving video monitoring. Related works use logic supervision to improve verification [8, 21, 43] and ensure safety in RL-based medical devices [44]. Isl advances the state of the art by introducing the first image segmentation logic for medical image analysis that unifies pixel and region-level reasoning, supports directional and spatial constraints, and incorporates model-derived attributes like confidence, with potential to improve both validation and learning.

7 Conclusion

In this paper, we present Isl, a novel specification language for expressing specifications in image segmentation. Isl supports qualitative and quantitative semantics, and we develop a monitoring algorithm to enable its practical application. Inspired by real-world medical image segmentation anatomical plausibility requirements, we evaluate the effectiveness of Isl on two medical imaging datasets. Results show that, even for well-performing models, Isl can accurately identify rare but critical segmentation failures. For more challenging segmentation tasks, the semantic insights provided by Isl offer an additional layer of safety and diagnostic value beyond traditional image segmentation metrics. Future work includes extending Isl to spatiotemporal settings, integrating it into learning frameworks for segmentation, and applying it to safety-critical domains like autonomous driving.

Acknowledgement

This material is based upon work supported by the National Science Foundation (NSF) under Award Numbers 2220401 and CNS-2443803. The paper reflects only the authors' view and does not necessarily reflect the views of the sponsoring agencies.

References

1. An, Z., Johnson, T.T., Ma, M.: Formal logic enabled personalized federated learning through property inference. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 10882–10890 (2024)
2. An, Z., Wang, X., Baier, H., Chen, Z., Dubey, A., Johnson, T.T., Sprinkle, J., Mukhopadhyay, A., Ma, M.: Combining llms with a logic-based framework to explain mcts. In: Proceedings of the 24th International Conference on Autonomous

Agents and Multiagent Systems. p. 2405–2407. AAMAS ’25, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2025)

3. An, Z., Wang, X., T. Johnson, T., Sprinkle, J., Ma, M.: Runtime monitoring of accidents in driving recordings with multi-type logic in empirical models. In: International Conference on Runtime Verification. pp. 376–388. Springer (2023)
4. Bennett, B.: Spatial reasoning with propositional logics. In: Principles of knowledge representation and reasoning. pp. 51–62. Elsevier (1994)
5. Bennett, B., Cohn, A.G., Wolter, F., Zakharyaschev, M.: Multi-dimensional modal logic as a framework for spatio-temporal reasoning. Applied Intelligence **17**, 239–251 (2002)
6. Berger, A.H., Lux, L., Stucki, N., Bürgin, V., Shit, S., Banaszak, A., Rueckert, D., Bauer, U., Paetzold, J.C.: Topologically faithful multi-class segmentation in medical images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 721–731. Springer (2024)
7. Bresolin, D., Sala, P., Della Monica, D., Montanari, A., Sciavicco, G.: A decidable spatial generalization of metric interval temporal logic. In: 2010 17th International Symposium on Temporal Representation and Reasoning. pp. 95–102. IEEE (2010)
8. Chang, K.K.C., Xu, K., Kim, E., Sangiovanni-Vincentelli, A., Seshia, S.A.: Dynamic, multi-objective specification and falsification of autonomous cps. In: International Conference on Runtime Verification. pp. 40–58. Springer (2024)
9. Chaudhry, R., Omole, A.E., Bordoni, B.: Anatomy, thorax, lungs. In: StatPearls [Internet]. StatPearls Publishing (2024)
10. Chen, Z., An, Z., Reynolds, J., Mullen, K., Martini, S., Ma, M.: Logidebrief: A signal-temporal logic based automated debriefing approach with large language models integration. arXiv preprint arXiv:2505.03985 (2025)
11. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945). https://doi.org/10.2307/1932409
12. Dokhanchi, A., Amor, H.B., Deshmukh, J.V., Fainekos, G.: Evaluating perception systems for autonomous vehicles using quality temporal logic. In: Runtime Verification: 18th International Conference, RV 2018, Limassol, Cyprus, November 10–13, 2018, Proceedings 18. pp. 409–416. Springer (2018)
13. European Union: Artificial intelligence act – article 15: Accuracy, robustness and cybersecurity. https://artificialintelligenceact.eu/article/15/ (2024), accessed: 2025-05-05
14. Gaggion, N., Mosquera, C., Mansilla, L., Saidman, J.M., Aineseder, M., Milone, D.H., Ferrante, E.: Chexmask: a large-scale dataset of anatomical segmentation masks for multi-center chest x-ray images. Scientific Data **11**(1), 511 (2024)
15. Gaggion, N., Mansilla, L., Mosquera, C., Milone, D.H., Ferrante, E.: Improving anatomical plausibility in medical image segmentation via hybrid graph neural networks: Applications to chest x-ray analysis. IEEE Transactions on Medical Imaging (2022)
16. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. International Conference on Machine Learning (ICML) (2016)
17. Gill, G., Bauer, C., Beichel, R.R.: A method for avoiding overlap of left and right lungs in shape model guided segmentation of lungs in ct volumes. Medical Physics **41**(10), 101908 (2014)
18. Gol, E.A., Bartocci, E., Belta, C.: A formal methods approach to pattern synthesis in reaction diffusion systems. In: 53rd IEEE Conference on Decision and Control. pp. 108–113. IEEE (2014)

19. Haghighi, I., Jones, A., Kong, Z., Bartocci, E., Gros, R., Belta, C.: Spatel: a novel spatial-temporal logic and its applications to networked systems. In: Proceedings of the 18th International Conference on Hybrid Systems: Computation and Control. pp. 189–198 (2015)
20. Hao, J., Shen, T., Zhu, X., Liu, Y., Behera, A., Zhang, D., Chen, B., Liu, J., Zhang, J., Zhao, Y.: Retinal structure detection in octa image via voting-based multitask learning. IEEE Transactions on Medical Imaging **41**(12), 3969–3980 (2022)
21. He, W., Wu, C., Bensalem, S.: Box-based monitor approach for out-of-distribution detection in yolo: An exploratory study. In: International Conference on Runtime Verification. pp. 229–239. Springer (2024)
22. Hekmatnejad, M., Hoxha, B., Deshmukh, J.V., Yang, Y., Fainekos, G.: Formalizing and evaluating requirements of perception systems for automated vehicles using spatio-temporal perception logic. The International Journal of Robotics Research **43**(2), 203–238 (2024)
23. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., et al.: Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 590–597 (2019)
24. Jaccard, P.: Étude comparative de la distribution florale dans une portion des alpes et des jura. Bull Soc Vaudoise Sci Nat **37**, 547–579 (1901)
25. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? Advances in Neural Information Processing Systems (NeurIPS) **30** (2017)
26. Kingma, D.P.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
27. Li, M., Huang, K., Xu, Q., Yang, J., Zhang, Y., Ji, Z., Xie, K., Yuan, S., Liu, Q., Chen, Q.: Octa-500: a retinal dataset for optical coherence tomography angiography study. Medical image analysis **93**, 103092 (2024)
28. Li, M., Huang, K., Xu, Q., Yang, J., Zhang, Y., Yang, J.: Octa-500: A novel dataset for retinal optical coherence tomography angiography analysis. Medical Image Analysis **91**, 103092 (2024). https://doi.org/10.1016/j.media.2024.103092, https://www.sciencedirect.com/science/article/pii/S1361841524000173
29. Li, T., Liu, J., Kang, J., Sun, H., Yin, W., Chen, X., Wang, H.: Stsl: A novel spatio-temporal specification language for cyber-physical systems. In: 2020 IEEE 20th International Conference on Software Quality, Reliability and Security (QRS). pp. 309–319. IEEE (2020)
30. Lim, J., Abily, A., Ben Salem, D., Gaillandre, L., Attye, A., Ognard, J.: Training and validation of a deep learning u-net architecture general model for automated segmentation of inner ear from ct. European Radiology Experimental **8**(1), 104 (2024)
31. Ma, M., Bartocci, E., Lifland, E., Stankovic, J., Feng, L.: Sastl: Spatial aggregation signal temporal logic for runtime monitoring in smart cities. In: 2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS). pp. 51–62. IEEE (2020)
32. Ma, M., Gao, J., Feng, L., Stankovic, J.: Stlnet: Signal temporal logic enforced multivariate recurrent neural networks. Advances in Neural Information Processing Systems **33**, 14604–14614 (2020)
33. Maler, O., Nickovic, D.: Monitoring temporal properties of continuous signals. In: International Symposium on Formal Techniques in Real-Time and Fault-Tolerant Systems. pp. 152–166. Springer (2004)

34. Moon, N., Bullitt, E., Van Leemput, K., Gerig, G.: Automatic brain and tumor segmentation. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002: 5th International Conference Tokyo, Japan, September 25–28, 2002 Proceedings, Part I 5. pp. 372–379. Springer (2002)
35. Nair, T., Precup, D., Arbel, T.: Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. Medical Image Analysis **59**, 101557 (2020)
36. Panara, K., Hoffer, M.: Anatomy, head and neck, ear internal auditory canal (internal auditory meatus, internal acoustic canal) (2019)
37. Randell, D.A., Cui, Z., Cohn, A.G.: A spatial logic based on regions and connection. KR **92**, 165–176 (1992)
38. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
39. Siddique, N., Paheding, S., Elkin, C.P., Devabhaktuni, V.: U-net and its variants for medical image segmentation: A review of theory and applications. IEEE access **9**, 82031–82057 (2021)
40. Sousa, A.M., Martins, S.B., Falcao, A.X., Reis, F., Bagatin, E., Irion, K.: Altis: A fast and automatic lung and trachea ct-image segmentation method. Medical physics **46**(11), 4970–4982 (2019)
41. Stucki, N., Paetzold, J.C., Shit, S., Menze, B., Bauer, U.: Topologically faithful image segmentation via induced matching of persistence barcodes. In: International Conference on Machine Learning. pp. 32698–32727. PMLR (2023)
42. Toledo, F., Woodlief, T., Elbaum, S., Dwyer, M.B.: Specifying and monitoring safe driving properties with scene graphs. In: 2024 IEEE International Conference on Robotics and Automation (ICRA). pp. 15577–15584. IEEE (2024)
43. Tran, H.D., Yang, X., Manzanas Lopez, D., Musau, P., Nguyen, L.V., Xiang, W., Bak, S., Johnson, T.T.: Nnv: the neural network verification tool for deep neural networks and learning-enabled cyber-physical systems. In: International conference on computer aided verification. pp. 3–17. Springer (2020)
44. Vuppala, S.R.H., Allen, N., Pinisetty, S., Roop, P.: A formal approach for safe reinforcement learning: A rate-adaptive pacemaker case study. In: International Conference on Runtime Verification. pp. 3–21. Springer (2024)
45. Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., Nandi, A.K.: Medical image segmentation using deep learning: A survey. IET image processing **16**(5), 1243–1267 (2022)
46. Wolter, F., Zakharyaschev, M.: Spatial representation and reasoning in rcc-8 with boolean region terms. In: Proceedings of the 14th European Conference on Artificial Intelligence. pp. 244–248. Citeseer (2000)
47. Woodlief, T., Toledo, F., Elbaum, S., Dwyer, M.B.: S3c: Spatial semantic scene coverage for autonomous vehicles. In: Proceedings of the IEEE/ACM 46th International Conference on Software Engineering. pp. 1–13 (2024)
48. Woodlief, T., Toledo, F., Elbaum, S., Dwyer, M.B.: The sgsm framework: Enabling the specification and monitor synthesis of safe driving properties through scene graphs. Science of Computer Programming **242**, 103252 (2025)
49. Zhao, L., Tohka, J., Ruotsalainen, U.: Accurate 3d left-right brain hemisphere segmentation in mr images based on shape bottlenecks and partial volume estimation. In: Image Analysis: 15th Scandinavian Conference, SCIA 2007, Aalborg, Denmark, June 10-14, 2007 15. pp. 581–590. Springer (2007)