Research Article

From Theory to Practice: A Case Study on EL-MIATTs Framework for Bicycle Lane Segmentation in Street Images

Yongquan Yang[1]

1. Institute of Sciences for AI, China

The EL-MIATTs (Evaluation and Learning with Multiple Inaccurate True Targets) framework is a novel scientific paradigm, developed based on the core principles of LAF (Logical Assessment Formula) and UTTL (Undefinable True Target Learning), aimed at tackling situations where acquiring the true target for a machine learning task is difficult, expensive, or fundamentally impossible. In this paper, bridging theory and practice, we present a case study applying the EL-MIATTs framework to the bicycle lane segmentation task in street images—a task chosen due to its inherent ambiguity and challenges in consistently defining and delineating bicycle lanes in complex urban scenes. We demonstrate that by adhering to the core principles of LAF and UTTL, an appropriate implementation of the EL-MIATTs framework can generate an optimized predictive model that uncovers the underlying true target for a given input street image, achieved through evaluation and learning that leverage the distributional representation of diverse properties encoded within the MIATTs corresponding to the true target underlying that input. Our results highlight the framework's potential to effectively handle scenarios where obtaining accurate true targets is challenging, offering a scientific and adaptable approach for real-world tasks. This case study thus validates the feasibility of translating the EL-MIATTs theoretical paradigm into practical applications, paving the way for further exploration and broader adoption in related domains.

Corresponding author: Yongquan Yang, remy_yang@foxmail.com or yongquan.yang@sciences4ai.com

1. Introduction

Operating under the relaxed yet shared assumption that the true target for a given machine learning task may not exist as a well-defined object in the real world, LAF [1] and UTTL [2] together provide a theoretical foundation for evaluating and learning predictive models without requiring a precise or universally agreed-upon ground truth target. Building on this foundation, the EL-MIATTs framework [3] has been proposed as a novel scientific paradigm [4] that, rather than relying on a single accurate true target, leverages multiple inaccurate true targets (MIATTs), each reflecting partial, noisy, or biased aspects of the latent true target. Through LAF-based evaluation and UTTL-guided learning, the framework iteratively improves model predictions, aiming to uncover an approximation of the latent true target by distilling useful signals from the distributional representation of its diverse properties, as encoded in the MIATTs for that target. This

makes EL-MIATTs particularly suited to scenarios where obtaining the true target for a machine learning task is costly, subjective, inconsistent, or fundamentally infeasible—such as medical diagnosis, aesthetic evaluation, or complex scene segmentation under ambiguous definitions.

In this paper, bridging theory and practice, we present a case study applying the EL-MIATTs framework to bicycle lane segmentation in street images—a task chosen for its inherent ambiguity and annotation challenges. Bicycle lane segmentation in street images, a representative problem within image semantic segmentation [5][6], is inherently challenging due to worn or faded markings, visual similarity with adjacent regions (e.g., sidewalks, bus lanes, parking areas), occlusions by vehicles or pedestrians, and lighting or weather variations that obscure lane boundaries [7]. Furthermore, design variability across cities and countries—in color, width, and symbols—complicates model generalization. Similar challenges have been reported in urban lane segmentation, where thin structures and low contrast lead to high misclassification rates [8]. These factors make the task prone to confusion and noise in annotations, providing a realistic and demanding benchmark for evaluating the EL-MIATTs framework, especially in bridging the gap from theory to practical deployment.

Section 2 presents the detailed formulations underpinning the EL-MIATTs framework, along with a description of the bicycle lane segmentation task in street images within the broader context of image semantic segmentation, and a discussion of its associated technical and practical challenges.

Specifically, we implemented an EL-MIATTs framework–based solution tailored to the bicycle lane segmentation task in street images. In this context, traditional supervised learning approaches—which depend on a single, accurate, and universally agreed-upon target—are often infeasible or prone to subjectivity and bias [7][8]. The EL-MIATTs framework addresses this limitation by introducing MIATTs, which approximate the latent true target as a distribution of diverse plausible interpretations. These MIATTs are represented in a structured and logically consistent manner, enabling the evaluation and learning processes to capture richer and more nuanced representations of the underlying target. By integrating MIATTs into both the evaluation and learning processes—grounded in the principles of LAF and UTTL [1][2]—the framework supports the development of predictive models that are inherently more robust to uncertainty and inconsistency in target definitions.

The practical implementation for bicycle lane segmentation follows a four-step pipeline: 1) MIATTs generation assisted by image segmentation tools; 2) Model construction using a U-Net architecture configured as the task-specific segmentation model; 3) Metric definition based on LAF [1] and conventional image semantic segmentation metrics [9] [10] to quantify discrepancies between the model's predicted true target and the MIATTs; and 4) Evaluation and learning for model optimization, guided by LAF-based metrics and the multi-target learning procedure of UTTL [2], producing a final optimized task-specific U-Net model that predicts a distribution over multiple plausible target properties from which the desired true target can be inferred. Further details on this specific EL-MIATTs implementation for bicycle lane segmentation in street images are provided in Section 3.

To validate the feasibility and effectiveness of the proposed EL-MIATTs framework, we designed and conducted a series of controlled experiments and corresponding analysis focusing on the bicycle lane segmentation task in street images.

The experimental workflow encompassed data preparation, hyperparameter settings for model training, intermediate result monitoring, and performance evaluation across multiple epochs using both quantitative and qualitative measures. A diverse set of LAF-based assessment metrics, including logical precision, recall, IoU, and error counts, were employed to capture different aspects of model performance and optimization convergence. Visualization of intermediate predictions and binary segmentations provided further insights into the evolving distribution of the underlying true target during training, while comparative analyses against MIATTs and logical true targets allowed for a deeper interpretation of the model's predictive behavior. This integrative approach not only demonstrates the technical soundness of the framework but also offers interpretability and transparency in understanding how the model progresses toward uncovering the latent true target. Section 4 provides further detailed experiments and analysis.

The EL-MIATTs framework, by explicitly modeling uncertainty and accommodating multiple plausible interpretations of a latent target, holds strong potential for applications where precise and universally accepted ground truth target annotations are infeasible. Such scenarios arise not only in street-scene understanding tasks, such as bicycle lane segmentation, but also in fields like medical imaging, remote sensing, and natural language understanding, where subjective judgment, incomplete information, and heterogeneous annotation standards challenge traditional supervised learning approaches. By integrating logically consistent but imperfect supervision into both evaluation and learning, the EL-MIATTs framework offers a robust and interpretable alternative suitable for high-stakes or resource-constrained contexts. Its adaptability enables scalable, domain-specific implementations that bridge theoretical rigor with practical deployment. More details on the broader implications of the EL-MIATTs framework derived in this paper are discussed in Section 5.

In summary, the contributions of this paper are as follows:

- We bridge theory and practice with a case study applying the EL-MIATTs framework to bicycle lane segmentation in street images.
- Building on LAF and UTTL principles, we specifically implement an EL-MIATTs-based solution tailored to this segmentation task.
- We design and conduct controlled experiments to validate the framework's feasibility and effectiveness.
- We demonstrate through analysis how MIATTs enable effective evaluation and learning for bicycle lane segmentation in undefinable true-target scenarios.
- We derive broader implications, showing EL-MIATTs' potential for tasks with uncertain or infeasible true target annotations by modeling uncertainty of the latent true target with a distribution of multiple inaccurate interpretations.

The remainder of this paper is structured as follows: Section 2 provides the theoretical foundations, practical applications, and associated challenges related to this work. Section 3 presents the details of the EL-MIATTs framework implementation for bicycle lane segmentation in street images. Section 4 illustrates the conducted experiments and corresponding analysis. Section 5 derives broader implications of the EL-MIATTs framework for future applications. Finally, Section 6 discusses the overall contributions, limitations, and avenues for future research.

2. Theoretical foundations, practical applications, and associated challenges

2.1. Formulations of EL-MIATTs Framework

Drawing upon the recent study [3], the essential connotation of MIATTs lies in their logical interpretation as structured approximations of the true target, constructed under conditions where the ground truth is undefinable. Under the assumption that the true target for a given machine learning task is not assumed to exist as a well-defined object in the real world, the connotation of MIATTs can be formalized as follows.

Connotation 1 (Multiple Inaccurate True Targets, MITTs). *Let t^* denote the (possibly undefinable) underlying true target corresponding to a given instance in a machine learning task. A set of MIATTs, denoted as $MIATTs = \{{t_n}^* | n \in \{1, \cdots, N\}\}$, is defined as a collection of target approximations such that:*

- *Each ${t_n}^* \in MIATTs$ is an inaccurate, partially informative surrogate of t^*;*
- *The collective set $MIATTs$ encodes a distribution over multiple plausible interpretations of the underlying true target;*
- *The goal is to leverage the diversity within $MIATTs$ to facilitate robust evaluation and learning when t^* is ambiguous, underspecified, or practically unobservable.*

This formulation enables evaluation and learning procedures that are not reliant on a single definitive ground truth, but rather on the distributional semantics of supervision across multiple weak signals.

On the basis of the connotation of MIATTs, the EL-MIATTs framework is formulated as follows.

$$\begin{gathered} MIATTs = GM\left(i, l; \theta^{GM}\right) = \{{t_n}^* | n \in \{1, \cdots, N\}\}, \\ N \geq 2 \;\&\; SF({t_n}^*) \subset SF(t^*) \;\&\; \bigcup_{n=1}^{N} SF({t_n}^*) \subseteq SF(t^*) \end{gathered} \tag{1}$$

$$\tilde{t} = PM(i; \theta^{PM}), \tag{2}$$

$$UTRs = D\left(\tilde{t}, MIATTs; \theta^{D}\right), \tag{3}$$

$$Ms = E\left(\tilde{t}, MIATTs, UTRs; \theta^{E}\right), \tag{4}$$

$$e = EP\left(\tilde{t}, MIATTs | Ms\right) = Ms(\tilde{t},\ MIATTs), \tag{5}$$

$$\begin{gathered} \tilde{\theta}^{PM} = LP\left(e = Ms\left(\tilde{t},\ MIATTs\right) | \theta^{LP}\right) \\ = arc\min_{\theta^{PM}} \left(Ms\left(\tilde{t} = PM\left(i; \theta^{PM}\right),\ MIATTs\right) | \theta^{LP}\right), \end{gathered} \tag{6}$$

$$\bar{t} = PM\left(i; \tilde{\theta}^{PM}\right). \tag{7}$$

Here, each θ in the above formulas respectively denotes the parameters for constructing each corresponding function.

Specifically, Formula (1) defines the function for generating the MIATTs (GM) from an instance i and its associated label l, which may be unavailable in practice. Formula (2) defines the predictive model (PM), which maps an input instance i to a predicted true target $\tilde{t}$. Formulas (3) and (4) respectively define the deduction function (D) for deriving underlying theoretical results (UTRs) from the predicted true target $\tilde{t}$ and the MIATTs, and the establishment function (E) for constructing computable evaluation metrics (Ms) based on the predicted true target $\tilde{t}$, MIATTs and the derived UTRs.

Formulas (5) and (6) then define the evaluation procedure (EP) and the learning procedure (LP) for the optimization of the PM in the context of MIATTs. Finally, Formula (7) defines the evolved PM with specified parameter $\tilde{\theta}^{PM}$, which is expected to map an instance to the desired underlying true target more accurately.

Together, these formulations form the backbone of the proposed theoretical EL-MIATTs framework for model evaluation and learning in settings characterized by multiple inaccurate approximations of the true target (MIATTs), offering a principled alternative where accurate true targets are unavailable or ill-defined. The EL-MIATTs framework, provided with an input instance and its possibly associated label, produces an optimized predictive model capable of predicting the desired underlying true target for that input, which is accomplished via evaluation and learning that utilize the distributional representation of diverse properties encoded within the MIATTs specific to that instance.

2.2. Image semantic segmentation

Image semantic segmentation is a fundamental task in computer vision that involves partitioning an image into semantically meaningful regions, typically to isolate objects or delineate boundaries between different classes [5][6]. The goal is to assign a class label to each pixel such that pixels with similar visual characteristics or spatial semantics are grouped together. This process is essential for downstream tasks such as object detection, scene understanding, medical image analysis, and autonomous driving.

Traditional segmentation techniques rely on edge detection, region growing, or clustering methods (e.g., k-means, graph cuts) [11]. However, recent advances in deep learning, particularly fully convolutional networks (FCNs), have significantly improved the accuracy and robustness of segmentation models [12]. Architectures like U-Net [13] and DeepLab [14] have become standard for various segmentation tasks, especially in biomedical imaging and natural scene analysis. For a comprehensive overview of recent deep learning-based advances in this task, readers are referred to [5].

2.3. Bicycle lane segmentation in street images

Bicycle lane segmentation in street images, which is a typical task within image semantic segmentation, refers to the task of automatically identifying and delineating bicycle lane areas from visual data, typically captured by cameras mounted on vehicles. It is a specialized form of semantic segmentation aimed at classifying each pixel in a street scene as part of a bicycle lane or not. This task is crucial for advanced driver-assistance systems (ADAS) and autonomous driving, as accurate bicycle lane detection enables safer navigation, supports cyclist-aware path planning, and helps ensure compliance with traffic regulations.

In the context of autonomous vehicles, robust bicycle lane segmentation contributes to improved environmental perception and context understanding, particularly in urban areas with mixed traffic. State-of-the-art methods can be leveraging deep convolutional neural networks (CNNs) or transformer-based architectures to achieve high segmentation accuracy [15][16].

2.4. Challenges of bicycle lane segmentation in street images

Depicting bicycle lane regions in street-scene images is a notably challenging task, due to several interrelated factors: 1) Variability and Degradation of Visual Features: Bicycle lane markings can be worn, faded, or partially obscured by debris, shadows, or seasonal effects like leaves and snow, leading to reduced visibility and segmentation ambiguity [7]; 2) Complex and Diverse Scene Layouts: Urban environments often feature overlapping or adjacent regions such as sidewalks, parking areas, or bus lanes, with visual characteristics similar to bicycle lanes. Additionally, bicycle lane styles (e.g., color, width, symbols) vary significantly across regions, making generalization difficult; 3) Illumination and Environmental Conditions: Variations in lighting—such as strong shadows, glare, or nighttime conditions—can obscure lane markings or distort their appearance; 4) Occlusion and Scene Clutter: Dynamic elements like vehicles, pedestrians, trees, and street furniture may partially or fully block the bicycle lane, complicating accurate pixel-level delineation. These issues collectively contribute to high within-class inconsistency and poor contrast with surrounding road surfaces, increasing the likelihood of misclassification in segmentation models.

Previous studies in related domains reflect similar difficulties: for example, aerial road marking segmentation faces challenges including abrasion, varying illumination, and occlusion by structures or vegetation [7]. Semantic segmentation for complex urban lane layouts also shows background misclassification and errors in thin lane structures [8]. Taken together, these findings underscore why bicycle lane segmentation in street scenes is an easily confused and technically demanding task. Our adoption of this task in validating the EL-MIATTs framework underscores not only its real-world relevance but also its methodological rigour in handling ambiguous and imperfect ground truth—precisely where our framework aims to bridge the gap from theory to practical application.

3. Implementation of EL-MIATTs framework for bicycle lane segmentation in street images

Following the formulations of the EL-MIATTs framework in Equations (1)-(7), this section details its implementation for bicycle lane segmentation in street images. Specifically, we cover image segmentation tool-assisted MIATTs generation, U-Net-based predictive model construction, LAF- and conventional metrics for image semantic segmentation (CMfISS)-driven metrics establishment, and metrics- and UTTL-driven procedures for evaluation and learning with MIATTs, concluding with a summary of these implementation details.

3.1. Image segmentation tool-assisted MIATTs generation

To generate the MIATTs (GM) from a given street image (which, in this case, lacks a given associated label), the parameters θ^{GM} must be specified to implement the GM function defined in Formula (1). Here, we follow a three-step process, which is aided by diverse image segmentation-related tools, to specify the parameters θ^{GM} for implementing Formula (1) in the context of bicycle lane segmentation in street images, which is visually summarized in Fig. 1. Further details are provided in the subsequent content of this subsection.

The first step involves conceptualizing the definition of bicycle lanes in street images. To support this process, we utilized the online demo of the Segment Anything Model (SAM) [17], which offers various interactive methods for object segmentation. Specifically, we used the "Everything" function of SAM to generate an overview of all segmented objects within a given street image. Subsequently, we employed the "Hover & Click" function, which allows users to add or remove segmentation masks by clicking, to gain more insight into how bicycle lanes might be delineated. While the segmentation results obtained using these two functions were not sufficient to definitively define bicycle lanes—particularly given our initial unfamiliarity with the task—they offered useful visual cues regarding their potential appearance and structure in street scenes. See the top row of Fig. 1 for illustration. While the segmentation results provide useful information, certain aspects remain ambiguous. For instance, it remains uncertain whether the bicycle symbol painted on the road surface or the two-dimensional shapes of people riding bicycles should be included within the mask that primarily represents the bicycle lane in the input street image. Based on our conceptual definition of a bicycle lane—assisted with the assistance of the SAM tool—we suggest that both the bicycle symbol and the two-dimensional representations of cyclists should be included as integral parts of the mask delineating the bicycle lane.

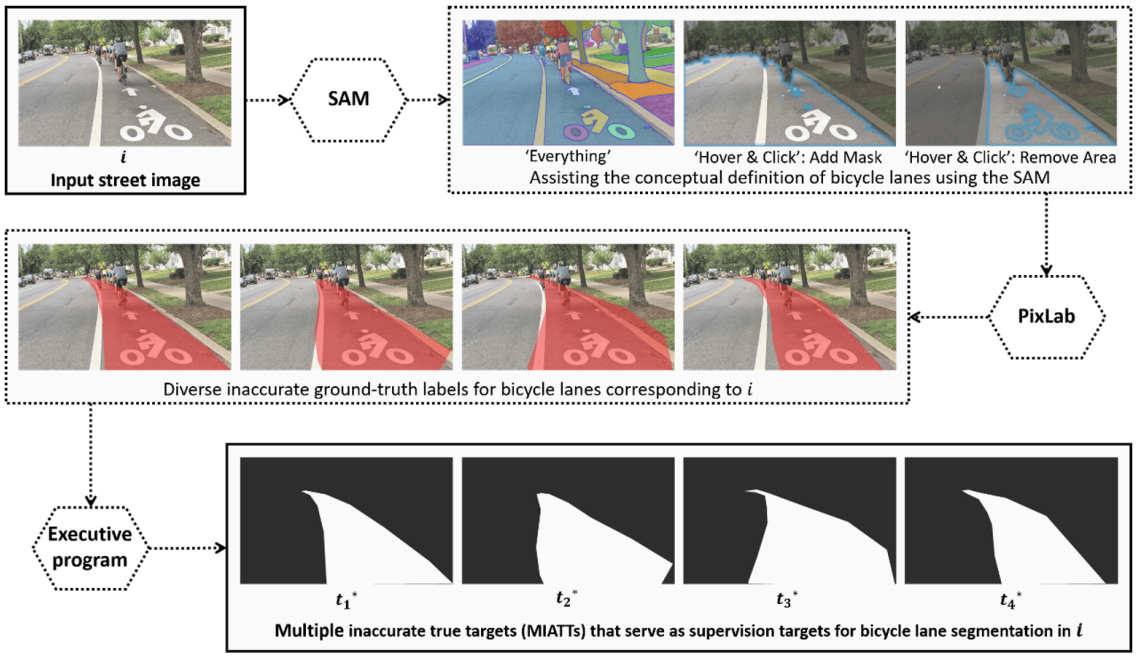

Figure 1. Three-step process illustrating the implementation of Formula (1) for generating MIATTs in the context of bicycle lane segmentation in street images. Top row: Step 1—conceptualization of bicycle lanes using SAM; Middle row: Step 2—annotation of inaccurate ground-truth labels using PixLab; Bottom row: Step 3—generation of MIATTs using a custom-developed executive program. Each pixel in an inaccurate true target is assigned a binary value: 0 (non-bicycle lane) or 1 (bicycle lane). To enhance visibility in the figure, the targets in the bottom row are visualized by scaling the values to 0 or 255. The input street image was photographed by Chris Roell and is used here with acknowledgment.

doi.org/10.32388/EZWLSN

The second step involves annotating various types of inaccurate ground-truth labels for identifying bicycle lanes in street images. Building upon our conceptual definition of a bicycle lane—formulated with the assistance of the SAM tool—we used the online annotation platform PixLab (https://pixlab.io) to generate four distinct types of inaccurate ground-truth labels. As illustrated in the middle row of Fig. 1, each label type partially covers the actual bicycle lane areas but also erroneously includes certain non-bicycle lane regions within the mask. These annotations reflect common forms of imprecision encountered in real-world labeling scenarios.

The third step focuses on generating MIATTs that serve as supervision targets for the task of bicycle lane segmentation in street images. Using the four distinct types of inaccurate ground-truth labels created with PixLab, we developed an executable program to synthesize MIATTs. As illustrated in the bottom row of Fig. 1, each individual inaccurate true target provides partial but meaningful information. Collectively, these MIATTs capture the essential characteristics of bicycle lanes, offering a more comprehensive representation than any single inaccurate true target alone. Notably, each pixel in an individual inaccurate true target is assigned a value of 0 (non-bicycle lane) or 1 (bicycle lane). For visualization purposes, each target shown in the bottom row of Fig. 1 is displayed by multiplying its original values by 255.

In summary, we utilized SAM, PixLab, and a custom-developed executive program to carry out the three steps described above. In other words, these tools were employed to specify the parameters in Formula (1) through their respective implicit functions. Table 1 provides detailed descriptions of each parameter and its associated implicit function.

Parameter	Implicit function
SAM	Using 'Everything' and 'Hover&Click' functions of SAM to assist the conceptual definition of bicycle lanes in street images
PixLab	Using PixLab to annotate diverse inaccurate ground-truth labels regarding the conceptual definition of bicycle lanes assisted with SAM
Executive program	Using custom-developed executive program to generate multiple inaccurate true targets (MIATTs) that serve as supervision targets for bicycle lane segmentation in street images, based on the diverse inaccurate ground-truth labels annotated with PixLab

Table 1. Descriptions for parameters and corresponding implicit functions used in the implementation of Formula (1).

Accordingly, based on Fig. 1 and Table 1, Formula (1), implemented with the specified parameters for generating MIATTs in the context of bicycle lane segmentation in street images, can be reformulated as follows.

$$\begin{aligned} MIATTs &= GM\left(i, l=\varnothing; \theta^{GM}=\{'SAM', 'PixLab', 'Executive\ program'\}\right) \\ &= Step3\left(Step2\left(Step1\left(i; 'SAM'\right); 'PixLab'\right); 'Executive\ program'\right) \\ &= \{{t_1}^*,\ {t_2}^*, {t_3}^*, {t_4}^*\},\ Val\left({t_n}^* \in MIATTs\right) \in \{0,1\} \end{aligned} \tag{8}$$

Here, $Step1$, $Step2$, and $Step3$ refer to the three-step process depicted in Fig. 1 for generating MIATTs, and $Val\left(t_n{}^*\right)$ indicates the pixel value within $t_n{}^* \in MIATTs$.

3.2. U-Net-based predictive model construction

To develop a predictive model (PM) for bicycle lane segmentation, the parameter θ^{PM} must be specified to implement the PM function defined in Formula (2). Here, we adopt the U-Net [13] architecture to specify the parameters θ^{PM} for implementing Formula (2), as predicting bicycle lane in street images can be considered a specific case of image semantic segmentation, for which U-Net is one of the most widely used deep neural network architectures [18][19]. Table 2 outlines the architecture of the U-Net model specifically developed for predicting bicycle lanes in street images.

Accordingly, based on Table 2, Formula (2), implemented with the specified parameter for predicting lane segmentation in street images, can be reformulated as follows.

$$\tilde{t} = PM\left(i; \theta^{PM} =' UNet^{'}\right) = UNet\left(i;\ \theta^{UNet}\right) \tag{9}$$

Here, $UNet$ denotes the task-specific predictive model designed using the U-Net architecture to address bicycle lane segmentation in street images, hereafter referred to as the *task-specific U-Net model*. Its architectural details are provided in Table 1. The symbol θ^{UNet} denotes the collection of learnable parameters subject to optimization during model learning.

Input and down sampling	**Up sampling and output**
Input image: Size(W, H,3)	Output prediction: ArgMax, Size(W, H,1)
DoubleConv(in_channels=3, out_channels=64): Size(W, H,64)	OutConv(in_channels=64, out_channels=2): Size(W, H,2)
Down(in_channels=64, out_channels=128): Size(W/2, H/2, 128)	Up(in_channels=128, out_channels=64): Size(W, H, 64)
Down(in_channels=128, out_channels=256): Size(W/4, H/4, 256)	Up(in_channels=256, out_channels=128): Size(W/2, H/2, 128)
Down(in_channels=256, out_channels=512): Size(W/8, H/8, 512)	Up(in_channels=512, out_channels=256): Size(W/4, H/4, 256)
Down(in_channels=512, out_channels=1024): Size(W/16, H/16, 1024)	Up(in_channels=1024, out_channels=512): Size(W/8, H/8, 512)

Table 2. Architecture details of the U-Net model developed for predicting bicycle lane in street images.

doi.org/10.32388/EZWLSN

3.3. LAF- and CMfISS-driven metrics establishment

We adopt logical assessment formula (LAF) [1] and conventional metrics for image semantic segmentation (CMfISS) [9] [10] to implement Equations (3) and (4) for appropriate metrics that can reasonably measure the discrepancy between predicted true target and MIATTs in the context of bicycle lane segmentation in street images.

Formulated in Equations (3) and (4), the construction of the metrics (Ms)—used to quantify the discrepancy between the predicted true target $\tilde{t}$ and its corresponding MIATTs for an input instance i—comprises two components: (1) Deducing underlying theoretical results (UTRs) regarding the predicted true target $\tilde{t}$ and the $MIATTs$ for the input instance i; and (2) Establishing computable equations for the Ms based on the predicted true target $\tilde{t}$, MIATTs and the deduced UTRs. Accordingly, the parameters θ^D and θ^E must be specified to implement the UTRs deduction (D) and computable equation establishment (E), two functions defined in Equations (3) and (4), respectively. As suggested in the recent study [3], LAF and task-specific evaluation metrics can be used to specify the parameters θ^D and θ^E for implementing Equations (3) and (4), respectively. Following this approach, the establishment details are provided in the rest contents of this subsection.

3.3.1. LAF-driven deduction of UTRs

Specifically, the principles of LAF and its validated applicability to image semantic segmentation evaluation, as demonstrated in prior studies [1][20], show that LAF, on the basis of MIATTs, can reasonably approximate conventional evaluation based on accurate true targets in challenging medical image semantic segmentation tasks. This enables effective measurement of the discrepancy between the predicted true target and its corresponding MIATTs for an input image. Building upon these prior works [1][20], we deduce the following UTR 1 in the context of bicycle lane segmentation in street images.

UTR 1. *LAF can approximate conventional accurate true target-based evaluation reasonably well in the task of bicycle lane segmentation in street images for measuring the discrepancy between the predicted true target $\tilde{t}$ of the task-specific U-Net model and its corresponding MIATTs for an input street image i.*

The primary rationale for deducing UTR1 lies in the substantial ambiguity and inconsistency observed in human annotations aided by diverse image segmentation tools when attempting to define bicycle lanes in street images, as discussed in Section 3.1. These inconsistencies reflect the subjective and context-dependent nature of the task, suggesting that the delineation of bicycle lanes is inherently prone to variation across annotators. Consequently, such characteristics justify the application of the LAF for evaluating segmentation quality in this domain. While some may contend that identifying bicycle lanes is occasionally straightforward, this opposing view only applies to a subset of cases and does not fundamentally weaken the general conclusion embodied in UTR1.

Building upon UTR1 and the formal interpretation of MIATTs (as outlined in Connotation 1), UTR2 can be logically deduced in the context of bicycle lane segmentation in street imagery as follows.

UTR 2. *Denote I is the space of the street image i. Here in the context of bicycle lane segmentation in street images, each $t_n{}^* \in MIATTs$ is a binary mask indicating a plausible bicycle lane region. Based on logical intersection and union operations, we formally derive three mutually exclusive semantic regions within I.*

- ***Confident bicycle lane region**:*

  $C_{bike} = \bigcap_n t_n{}^* \in MIATTs$

  This region includes all pixels that are consistently marked as bicycle lane across all MIATTs and is thus interpreted as a high-confidence bicycle lane area.

- ***Confident non-bicycle lane region**:*

  $C_{non} = I \setminus (\bigcup_n t_n{}^* \in MIATTs)$

  This region includes all pixels that are consistently excluded by all MIATTs, and is thus interpreted as a high-confidence non-bicycle lane area.

- ***Uncertain region**:*

  $C_{uncertain} = I \setminus (C_{bike} \cup C_{non})$

  This region comprises the remaining pixels where the MIATTs disagree, and it is logically interpreted as uncertain—it may correspond to either bicycle lane or non-bicycle lane.

As a result, the entire street image space I is logically partitioned as:

$$I = C_{bike} \cup C_{non} \cup C_{uncertain},$$

$$C_{bike} \cap C_{non} = C_{bike} \cap C_{uncertain} = C_{non} \cap C_{uncertain} = \varnothing$$

This formulation provides a principled basis for distinguishing between high-confidence and uncertain regions when evaluating predicted true targets.

Accordingly, based on UTRs 1 and 2, Formula (3), implemented with the specified parameter for deducing UTRs, can be reformulated as follows.

$$\begin{aligned} UTRs &= D\left(\tilde{t}, MIATTs; \theta^D =' LAF'\right) \\ &= LAF\left(\tilde{t}, MIATTs\right) = \{UTR1,\ UTR2\} \end{aligned} \tag{10}$$

3.3.2. CMfISS-driven establishment of metrics based on UTRs

Referring to the deduced UTRs 1 and 2, the set of MIATTs, denoted as $MIATTs = \{t_1{}^*,\ t_2{}^*, t_3{}^*, t_4{}^*\}$, generated by Formula (8), can be logically transformed into a *logical true target* (LTT) that approximates certain properties of the underlying (yet undefinable) true target for the street image iii. This transformation process is illustrated in Fig. 2. In the resulting logical true target, the region C_{bike} is interpreted as the *logical object* class (i.e., bicycle lane), C_{non} as the *logical non-object* class (i.e., non-bicycle lane), and $C_{uncertain}$ as the *logical indeterminate* class, representing areas that may be either object or non-object. The formal representation of this transformation is given as follows.

$$LTT(MIATTs) = UTR1\&2(MIATTs) = \{C_{bike}, C_{non}, C_{uncertain}\}. \tag{11}$$

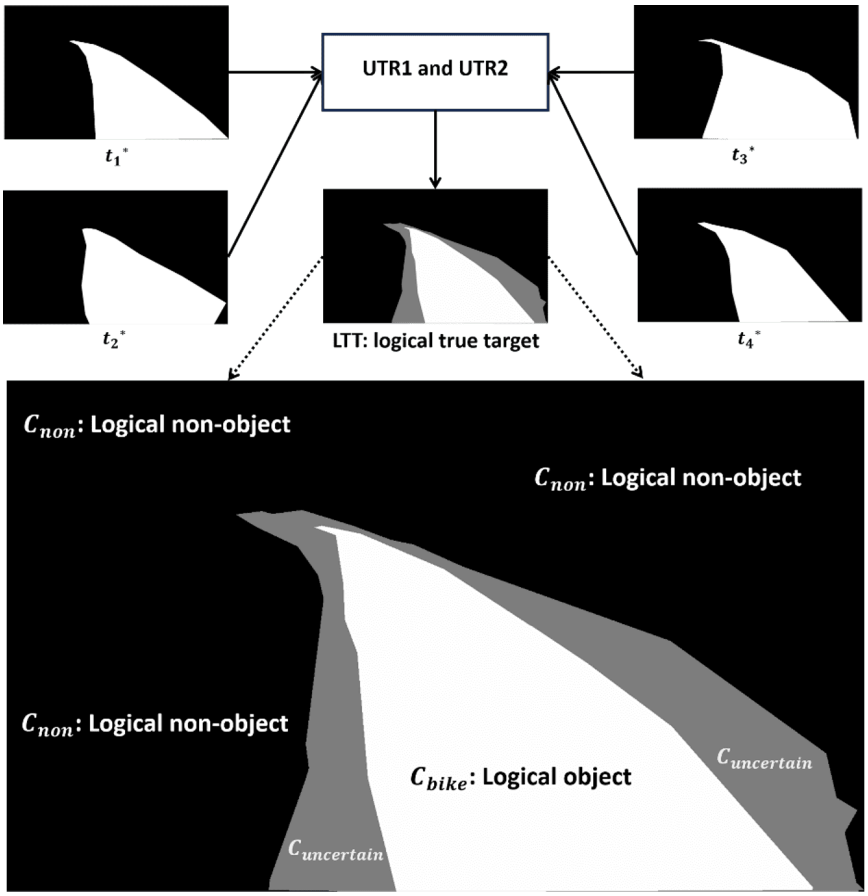

Figure 2. Visual representation of the logical transformation from MIATTs to a logical true target, as derived from UTRs 1 and 2.

Based on the LTT derived from the set of MIATTs, as formally defined in Formula (10), we proceed to formulate computable equations for quantifying the discrepancy between a predicted true target and the MIATTs. These formulations are informed by existing CMfISS. Consequently, we develop a new set of LAF-based metrics for image semantic segmentation, referred to as LAF-MfISS. The corresponding computable equations for both CMfISS and LAF-MfISS are comparatively presented in Table 3.

doi.org/10.32388/EZWLSN

Conventional metrics for image semantic segmentation (CMfISS)		LAF-based metrics for image semantic segmentation (LAF-MfISS)	
True positive (TP)	A pixel targeted as 'Object' is correctly predicted as 'Object'	**Logical TP (LTP)**	A pixel targeted as 'Logical object' is correctly predicted as 'Logical object'
False positive (FP)	A pixel targeted as 'Non-object' is incorrectly predicted as 'Object'	**Logical FP (LFP)**	A pixel targeted as 'Logical non-object' is incorrectly predicted as 'Logical object'
True negative (TN)	A pixel targeted as 'Non-object' is correctly predicted as 'Non-object'	**Logical TN (LTN)**	A pixel targeted as 'Logical non-object' is correctly predicted as 'Logical non-object'
False negative (FN)	A pixel targeted as 'Object' is incorrectly predicted as 'Non-object'	**Logical FN (LFN)**	A pixel targeted as 'Logical object' is incorrectly predicted as 'Logical non-object'
Errors	FP + FN	**Logical Errors (LErrors)**	LFP + LFN
Precision	TP / (TP+FP)	**Logical Precision (LPrecision)**	LTP / (LTP+LFP)
Recall	TP / (TP+FN)	**Logical Recall (LRecall)**	LTP / (LTP+LFN)
F1	2(Precision*Recall) / (Precision+Recall)	**Logical F1 (LF1)**	2(LPrecision*LRecall) / (LPrecision+LRecall)
Accuracy	(TP+TN) / (TP+FP+TN+FN)	**Logical Accuracy (LAccuracy)**	(LTP+LTN) / (LTP+LFP+LTN+LFN)
Intersection of union (IoU)	TP / (TP+FP+FN)	**Logical IoU (LIoU)**	LTP / (LTP+LFP+LFN)

Table 3. Comparison of CMfISS and LAF-MfISS through their respective definitions and computable equations.

Building upon the intermediate results—specifically Formula (10), Figure 2, and Table 3—we now reformulate Formula (4), incorporating specified parameter, to establish refined metrics for measuring the discrepancy between the predicted true target and the MIATTs.

$$\begin{aligned} Ms &= E\left(\tilde{t}, MIATTs, UTRs; \theta^{E} =' CMfISS'\right) \\ &= CMfISS\left(\tilde{t}, UTRs(MIATTs)\right) \\ &= CMfISS\left(\tilde{t}, UTR1\&2(MIATTs)\right) \\ &= \left\{ \begin{matrix} LTP,\ LFP, LTN, LFN, LErros \\ LPrecision,\ LRecall, LF1, LAccuracy,\ LIoU \end{matrix} \left(\tilde{t}, LTT(MIATTs)\right) \right\} \end{aligned} \tag{12}$$

3.4. Metrics- and UTTL-driven procedures for evaluation and learning with MIATTs

As formulated in Equation (5), the evaluation procedure (EP) with MIATTs operates by applying suitable metrics to quantify the discrepancy between the predicted true target and the MIATTs. In this study, we directly select the appropriate metrics from those defined in Equation (12) to implement the EP for the task of bicycle lane segmentation in street images.

As formulated in Equation (6), the learning procedure (LP) with MIATTs is inherently dependent on the EP, since its objective is to optimize the internal parameters of the predictive model by minimizing the EP's output. Therefore, the design and correctness of the LP are contingent on the proper implementation of the EP. A critical step in executing the EP within the LP framework is the specification of the parameter θ^{LP}. As suggested in recent work [3] on undefinable true target learning (UTTL) [2], which demonstrates the feasibility of learning from MIATTs, UTTL can be employed to determine θ^{LP} for Equation (6). Following this approach, we adopt UTTL to implement the LP for bicycle lane segmentation in street images.

3.5. Metrics-driven procedure for evaluation with MIATTs

All the metrics defined in Equation (12) are applicable to implementing the EP for bicycle lane segmentation. Accordingly, in the context of bicycle lane segmentation in street images, Equation (5) can be reformulated as follows.

$$\begin{aligned} e = EP\left(\tilde{t}, MIATTs | Ms\right) = Ms\left(\tilde{t}, MIATTs\right) \\ = \left\{ \begin{matrix} LTP\left(\tilde{t}, LTT(MIATTs)\right), \ldots, LIoU\left(\tilde{t}, LTT(MIATTs)\right) \\ 1 - LIoU\left(\tilde{t}, LTT(MIATTs)\right) \end{matrix} \right\}. \end{aligned} \tag{13}$$

To facilitate the subsequent implementation of the LP, in addition to the metrics defined in Equation (12), we also employ the evaluation metric $1 - LIoU\left(\tilde{t}, LTT(MIATTs)\right)$ for the EP, as described in Formula (13). The parent metric IoU of the LAF-derived metric LIoU in Formula (12) is widely used to assess the quality of predicted true targets in image semantic segmentation tasks, and its complement, 1-IoU, can naturally serve as a loss function [21][22][23]. This property enables a seamless transition from EP evaluation to the LP loss function.

3.4.2. UTTL-driven procedure for learning with MIATTs

On one hand, supported by the EP defined in Formula (13), we adopt $e = 1 - LIoU\left(\tilde{t}, LTT(MIATTs)\right)$ as the loss function in Equation (6) for implementing the LP in bicycle lane segmentation from street images. On the other hand, the principles of UTTL and its demonstrated applicability to image semantic segmentation tasks [2][24][25][26] indicate that UTTL can be practically realized through a multi-target learning (MTL) procedure based on MIATTs. With an appropriate loss function defined between the predicted true target and MIATTs, such a UTTL-driven MTL (UTTL-MTL) framework can effectively learn predictive models for challenging medical image semantic segmentation tasks [24][25][26]. Building upon these prior studies [2][24][25][26], we employ UTTL-MTL to determine θ^{LP} in Equation (6) for implementing the LP in bicycle lane segmentation from street images.

To deploy the UTTL-MTL procedure in practice, we select an appropriate optimization algorithm to evolve the task-specific U-Net model so as to minimize the loss function $e = 1 - LIoU\left(\tilde{t}, LTT(MIATTs)\right)$ for improved predicted true targets $\tilde{t}$. As the task-specific U-Net model is a typical deep neural network [18][19], it can be optimized using a stochastic gradient descent (SGD) algorithm [27][28] with respect to this loss. In particular, we adopt the widely used RMSprop variant of SGD (SGD-RMSprop) [29] to update the internal parameters of the U-Net model.

Accordingly, based on the loss function $e = 1 - LIoU\left(\tilde{t}, LTT(MIATTs)\right)$, UTTL-MTL, and SGD-RMSprop, Equation (6), implemented with the specified parameters for the EP in the context of bicycle lane segmentation in street images, can be reformulated as follows.

$$\begin{aligned}\tilde{\theta}^{UNet} &= LP\left(e = 1 - LIoU\left(\tilde{t},\ LTT(MIATTs)\right)|\theta^{LP} = \{'UTTLMTL',' SGDRMSpro'\}\right)\\ &= arc\min_{\theta^{PM}}\left(1 - LIoU\left(\tilde{t} = UNet\left(i;\theta^{UNet}\right),\ LTT(MIATTs)\right)|\left\{'UTTLMTL',' SGDRMSprop'\right\}\right)\\ &= UTTLMTL\left(1 - LIoU\left(UNet\left(i;\theta^{UNet}\right), LTT(MIATTs)\right)|opt = SGDRMSprop\right).\end{aligned} \tag{14}$$

In practice, directly using Formula (14) would cause the UTTL-MTL procedure to lose its MTL property, as $LTT(MIATTs)$ transforms MIATTs into a single logical true target partitioned into three parts: $\{C_{bike}, C_{non}, C_{uncertain}\}$. As suggested by previous studies [24][25][26], the UTTL-MTL procedure should learn directly from MIATTs to preserve its MTL property in image semantic segmentation. Therefore, we implemented an alternative loss function in Formula (14) that combines cross-entropy loss [30][31] and Dice loss [32][33] for the UTTL-MTL procedure, as follows.

$$\begin{aligned}L_{dice+CE}\left(\tilde{t}, MIATTs\right) &= \frac{1}{2}\left(dice\left(\tilde{t}, MIATTs\right) + CE\left(\tilde{t}, MIATTs\right)\right)\\ = \frac{1}{2}\left(\frac{1}{4}\sum_{i=1}^{4}\left(1 - \sum_{n=1}^{4}\frac{\tilde{t} * {t_n}^*}{\tilde{t} + {t_n}^*}\right)\right.&\left. - \frac{1}{4}\sum_{n=1}^{4}\left[{t_n}^* * \log\left(\tilde{t}\right) + \left(1 - {t_n}^*\right) * \log\left(1 - \tilde{t}\right)\right]\right).\end{aligned} \tag{15}$$

Accordingly, on the basis of Equation (15), we rewrite Equation (14) as follows.

$$\tilde{\theta}^{UNet} = UTTLMTL\left(L_{dice+CE}\left(\tilde{t} = UNet\left(i;\theta^{UNet}\right), MIATTs\right)|SGDRMSprop\right). \tag{16}$$

By using Equation (16) to minimize the loss $L_{dice+CE}\left(\tilde{t}, MIATTs\right)$, we obtain a series of optimized parameters $\left\{\tilde{\theta}_1^{UNet},\ \ldots, \tilde{\theta}_S^{UNet}\right\}$ for the task-specific U-Net models across different epochs. We then evaluate these models using $LIoU\left(\tilde{t}, LTT(MIATTs)\right)$ and select the one with the best performance for predicting the desired underlying true target. This process can be denoted as follows.

$$\bar{\theta}^{UNet} = arc\max_{\tilde{\theta}^{UNet}}\left(LIoU\left(\tilde{t} = UNet\left(i;\tilde{\theta}^{UNet}\right), LTT(MIATTs)\right)|\left\{\tilde{\theta}_1^{UNet},\ \ldots, \tilde{\theta}_S^{UNet}\right\}\right). \tag{17}$$

3.5. Summary

The implementation of the EL-MIATTs framework for bicycle lane segmentation in street images is outlined in Fig. 3.

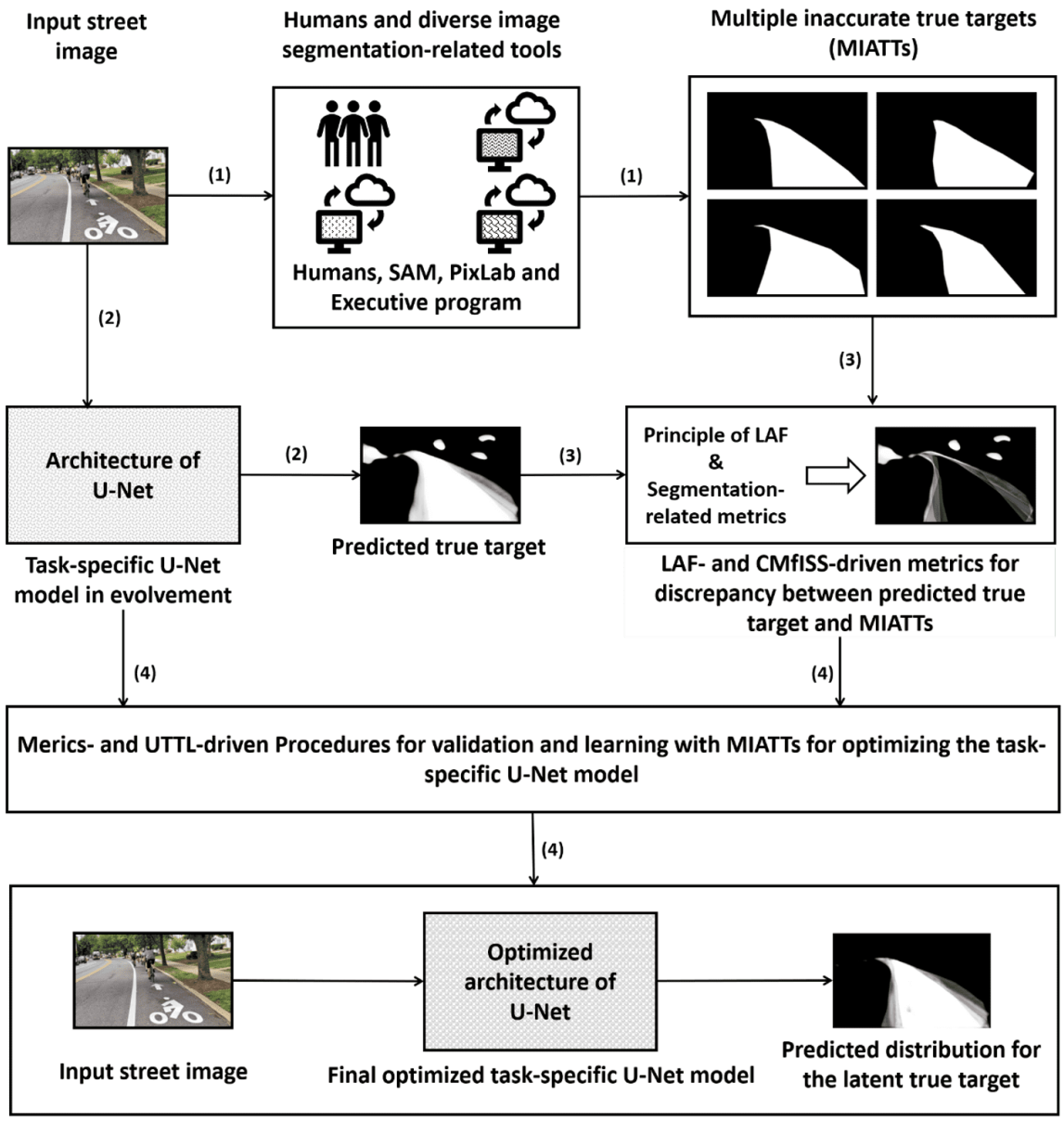

Figure 3. Visual summary of the EL-MIATTs framework implementation for the bicycle lane segmentation task on street images. The input street image was photographed by Chris Roell and is used here with acknowledgment.

The process comprises four main stages:

- **MIATTs Generation**: For an input street image, with the assistance of image segmentation-related tools, MIATTs are produced to encode a distribution over multiple plausible interpretations of the underlying true target.
- **Model Construction**: A U-Net-based predictive architecture is designed and configured as the task-specific segmentation model.
- **Metric Definition**: Segmentation evaluation metrics are established based on the LAF and CMfISS to quantify discrepancies between the model's predicted true target and the MIATTs.

- **Evaluation & Learning for Model Optimization**: Guided by the established LAF-based metrics and the multi-target learning procedure within UTTL, the model undergoes iterative evaluation and learning using the MIATTs, culminating in the selection of the final optimized task-specific U-Net model that is able to predict the distribution over multiple plausible properties of the underlying true target, from which the desired true target can be derived.

Finally, using the parameter set $\bar{\theta}^{UNet}$ obtained through Formulas (16) and (17), we derive the final evolved task-specific U-Net model capable of mapping an input street image iii to the desired true target $\bar{t}$. This mapping can be formally expressed as follows.

$$\bar{t} = UNet\left(i; \bar{\theta}^{UNet}\right). \tag{18}$$

4. Experiments and analysis

The experiments were designed as a proof-of-concept demonstration to verify the feasibility of implementing the EL-MIATTs framework in the context of bicycle lane segmentation. Rather than pursuing state-of-the-art results for a specific application, this focus is on validating the implementation workflow of the EL-MIATTs framework—from theory to practice—to substantiate its core theoretical claim: the EL-MIATTs framework is designed to produce an optimized predictive model that uncovers the underlying true target for a given input instance through evaluation and learning, leveraging the distributional representation of diverse properties encoded in the MIATTs specific to that instance.

This section is organized into four main parts. First, the experimental design subsection outlines the data preparation process, hyperparameter settings, and code implementation and computing platform adopted for this study. Next, the conducted experiments subsection details how the experiments were conducted following the experimental design. The experimental results subsection then presents both quantitative and qualitative outcomes, supported by relevant figures and tables. Finally, the analysis subsection provides an in-depth interpretation of the results, discussing performance trends, convergence behavior, model selection rationale, and the reasonability and interpretability of the selected model's prediction.

4.1. Experimental design

The experimental design comprises data preparation, hyperparameter settings for producing the final optimized task-specific U-Net model, and specifications of the code implementation and computing platform. These details are presented to facilitate reproducibility of the results and to clarify the computational resources utilized in the experiments.

4.1.1. Data preparation

Due to the objective of this study being methodological verification rather than dataset-specific performance benchmarking, a single representative sample was selected to minimize the annotation cost while still allowing for direct validation of the implementation feasibility.

The street image used in the experiments is shown as the “Input street image” in Fig. 3 (Section 3), serving to illustrate the implementation of the EL-MIATTs framework for bicycle lane segmentation in street scenes. The original resolution of the input street image is $952 \times 1600 \times 3$ (HWC). We scaled it by a factor of 0.5, resulting in a reshaped image of $476 \times 800 \times 3$. This reshaped image was then processed using Formula (8), producing four inaccurate true targets, which are presented as the “MIATTs” in Fig. 3.

Together, the reshaped input street image and its corresponding MIATTs constitute the data basis for the subsequent evaluation and learning procedures aimed at optimizing the task-specific U-Net, enabling it to appropriately predict the bicycle lane segmentation underlying the input street image.

4.1.2. Hyperparameter settings for producing the final optimized task-specific U-Net model

The task-specific U-Net model is optimized using the UTTL-MTL procedure, employing the RMSprop variant of the SGD optimizer, as implemented in Formula (16). The initial hyperparameter settings for RMSprop are: learning rate = 1e-5, weight decay =1e-8, and momentum = 0.999.

During optimization, following Formula (17) and based on the prepared dataset, the model is evaluated at each epoch using the evaluation procedure described in Formula (13). The stopping criteria are defined as LIoU>0.999 and LErrors<100. Once the first stopping criterion is met, learning continues for an additional four epochs to check whether a model with improved evaluation performance can be obtained. If such an improvement is found, learning proceeds; otherwise, the process terminates, and the final optimized task-specific U-Net model is selected at the epoch where the stopping criterion was first satisfied.

4.1.3. Specifications of the code implementation and computing platform

Our experimental code is implemented in Python using the PyTorch deep learning library. The computing platform used for running the experiments consists of a single GPU (RTX 4090D, 24 GB), a 16-core CPU (Xeon® Platinum), 80 GB of RAM, CUDA 12.4, PyTorch 2.0, Python 3.8, and the Ubuntu 20.04 operating system.

4.2. Conducted experiments

Following the experimental design, we carried out a series of experiments. Using the data prepared in Section 4.1.1 and the hyperparameter settings described in Section 4.1.2, we optimized the task-specific U-Net model with 700 epochs, saving model checkpoints every 20 epochs. This resulted in 35 task-specific U-Net models with different optimized parameters. According to the stopping criterion in Section 4.1.2, the model with the best predictive performance was obtained at epoch 620. We then performed statistical experiments to assess and compare the performance of the saved models. Additional details of these results and corresponding analysis are provided in the following subsection.

4.3. Experimental results

The experiments produced results comprising: (i) quantitative evaluations using the diverse metrics defined in Formula (13); (ii) selection of the final optimized task-specific U-Net model based on these evaluations; (iii) visualization of

intermediate outputs related to the task-specific U-Net model during optimization; and (iv) qualitative comparison between the underlying bicycle lane segmentation uncovered by the final optimized model and the MIATTs corresponding to the input street image.

4.3.1. Quantitative validation: Diverse metrics across epochs

Based on the evaluation metrics outlined in Formula (13) and the loss function defined in Formula (15), the quantitative performance of the 35 saved task-specific U-Net models—each representing a distinct set of optimized parameters obtained at different training epochs—is comprehensively summarized in Table 4. This tabulated comparison enables an objective assessment of model progression over the course of optimization, highlighting trends in predictive LAccuracy, LPrecision, LRecall, LF1, LIoU, and etc, and Loss reduction.

Epoch	LTP	LTN	LFP	LFN	LPrecision	LRecall	LF1	LAccuracy	LIoU	Loss
20	332516	70438	930668	0	0.263	1	0.417	0.302	0.263	0.711
40	332516	602333	398773	0	0.455	1	0.625	0.701	0.455	0.543
60	332516	734186	266920	0	0.555	1	0.714	0.800	0.555	0.466
80	332516	697802	303304	0	0.523	1	0.687	0.773	0.523	0.487
100	332516	726861	274245	0	0.548	1	0.708	0.794	0.548	0.468
120	332516	797843	203263	0	0.621	1	0.766	0.848	0.621	0.416
140	332516	771050	230056	0	0.591	1	0.743	0.827	0.591	0.437
160	332516	823053	178053	0	0.651	1	0.789	0.866	0.651	0.396
180	332516	878543	122563	0	0.731	1	0.844	0.908	0.731	0.342
200	332516	905012	96094	0	0.776	1	0.874	0.928	0.776	0.315
220	332516	924961	76145	0	0.814	1	0.897	0.943	0.814	0.291
240	332516	925335	75771	0	0.814	1	0.898	0.943	0.814	0.292
260	332516	931754	69352	0	0.827	1	0.906	0.948	0.827	0.284
280	332516	960246	40860	0	0.891	1	0.942	0.969	0.891	0.247
300	332516	979540	21566	0	0.939	1	0.969	0.984	0.939	0.220
320	332516	988748	12358	0	0.964	1	0.982	0.991	0.964	0.204
340	332516	985493	15613	0	0.955	1	0.977	0.988	0.955	0.212
360	332516	986447	14659	0	0.958	1	0.978	0.989	0.958	0.207
380	332516	996632	4474	0	0.987	1	0.993	0.997	0.987	0.192
400	332516	1000047	1059	0	0.997	1	0.998	0.999	0.997	0.182
420	332516	1000576	530	0	0.998	1	0.999	1.000	0.998	0.178
440	332516	1000597	509	0	0.998	1	0.999	1.000	0.998	0.181
460	332516	1000743	363	0	0.999	1	0.999	1.000	0.999	0.179
480	332516	1000489	617	0	0.998	1	0.999	1.000	0.998	0.183
500	332516	1000811	295	0	0.999	1	1.000	1.000	0.999	0.178
520	332516	1000784	322	0	0.999	1	1.000	1.000	0.999	0.177
540	332516	1000813	293	0	0.999	1	1.000	1.000	0.999	0.179
560	332516	1000953	153	0	1.000	1	1.000	1.000	1.000	0.174

doi.org/10.32388/EZWLSN

Epoch	LTP	LTN	LFP	LFN	LPrecision	LRecall	LF1	LAccuracy	LIoU	Loss
580	332516	1000811	295	0	0.999	1	1.000	1.000	0.999	0.178
600	332516	1000973	133	0	1.000	1	1.000	1.000	1.000	0.176
620	332516	1001053	53	0	1.000	1	1.000	1.000	1.000	0.173
640	332516	1000976	130	0	1.000	1	1.000	1.000	1.000	0.172
660	332516	1000924	182	0	0.999	1	1.000	1.000	0.996	0.179
680	332516	1000931	175	0	0.999	1	1.000	1.000	0.999	0.175
700	332516	1000959	147	0	1.000	1	1.000	1.000	1.000	0.173

Table 4. Quantitative performance comparison of the task-specific U-Net models at different epochs across various evaluation metrics defined in Formulas (13) and (15).

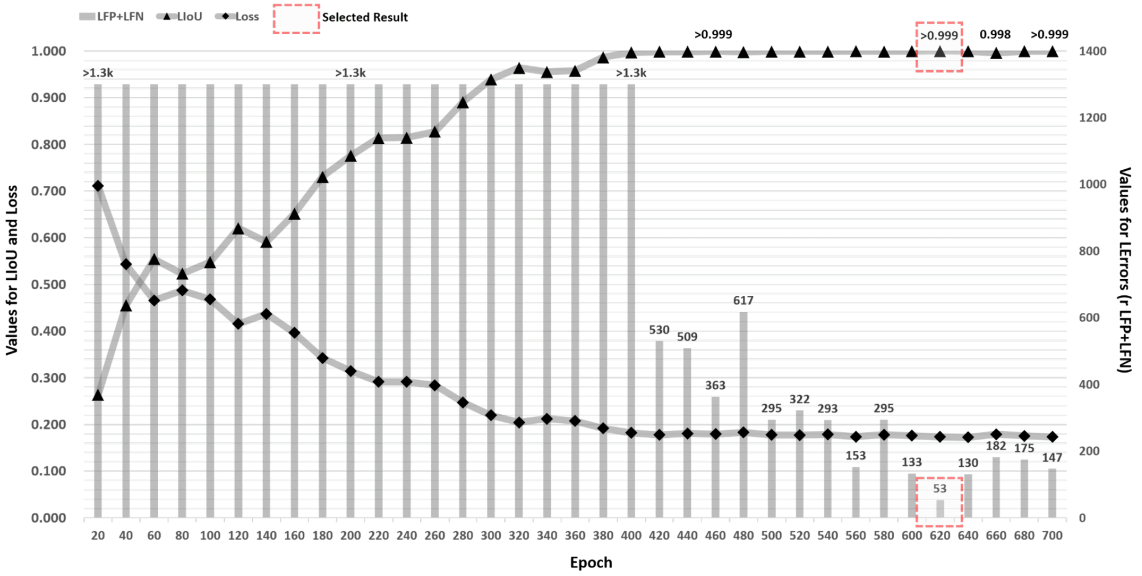

Figure 4. Evolution of LIoU, LErrors (LFP+LFN) and Loss over optimizing epochs. Values included by red bounding boxes represent the best performance across all epochs. Evolution of key performance indicators across learning epochs for the task-specific U-Net model, including LIoU (black triangles), Loss (black circles), and total LErrors (LFP+LFN, grey bars). The stop criteria (LIoU > 0.999 and LErrors < 100) were first satisfied at epoch 620 (highlighted), yielding LIoU = 1.0 and only 53 errors, which was selected as the final optimized model. Early epochs show rapid gains in LIoU and steep reductions in Loss, while errors remain high until mid-learning. After ~500 epochs, LIoU plateaus near perfection, but errors continue to drop until the optimal point at epoch 620, after which slight fluctuations indicate no further improvement.

doi.org/10.32388/EZWLSN

4.3.2. Selection of the final optimized task-specific U-Net model

To identify the final optimized task-specific U-Net model, we examined the temporal evolution of several key performance indicators—namely logical intersection-over-union (LIoU), logical error counts (LErrors, computed as LFP + LFN), and loss—across the optimization epochs, as depicted in Fig. 4. This visualization provides an intuitive overview of the model's learning dynamics, highlighting both the rapid performance gains in the early and mid-stages of optimizing and the gradual stabilization of metrics as convergence is approached. The trends in Fig. 4 also serve to pinpoint the epoch at which the model meets the predefined stopping criteria (LIoU > 0.999 and LErrors < 100), ensuring the selected model achieves both high accuracy and low error rates while avoiding unnecessary overtraining

4.3.3. Visualization of the task-specific U-Net model's intermediate results during the optimization across epochs

For the given input street image, Figures 5, 6, and 7 respectively visualize key intermediate results generated by the task-specific U-Net model throughout the optimization process across epochs. Specifically, Figure 5 presents the raw segmentation probability predictions, Figure 6 shows their corresponding binary segmentations obtained by applying a 0.5 threshold, and Figure 7 compares these binary masks against the logical true target (LTT) defined in Formula (10). Together, these visualizations provide a progressive view of how the model's predictions evolve during optimizing, from initial coarse outputs to refined segmentations that increasingly conform to the constraints of the LTT.

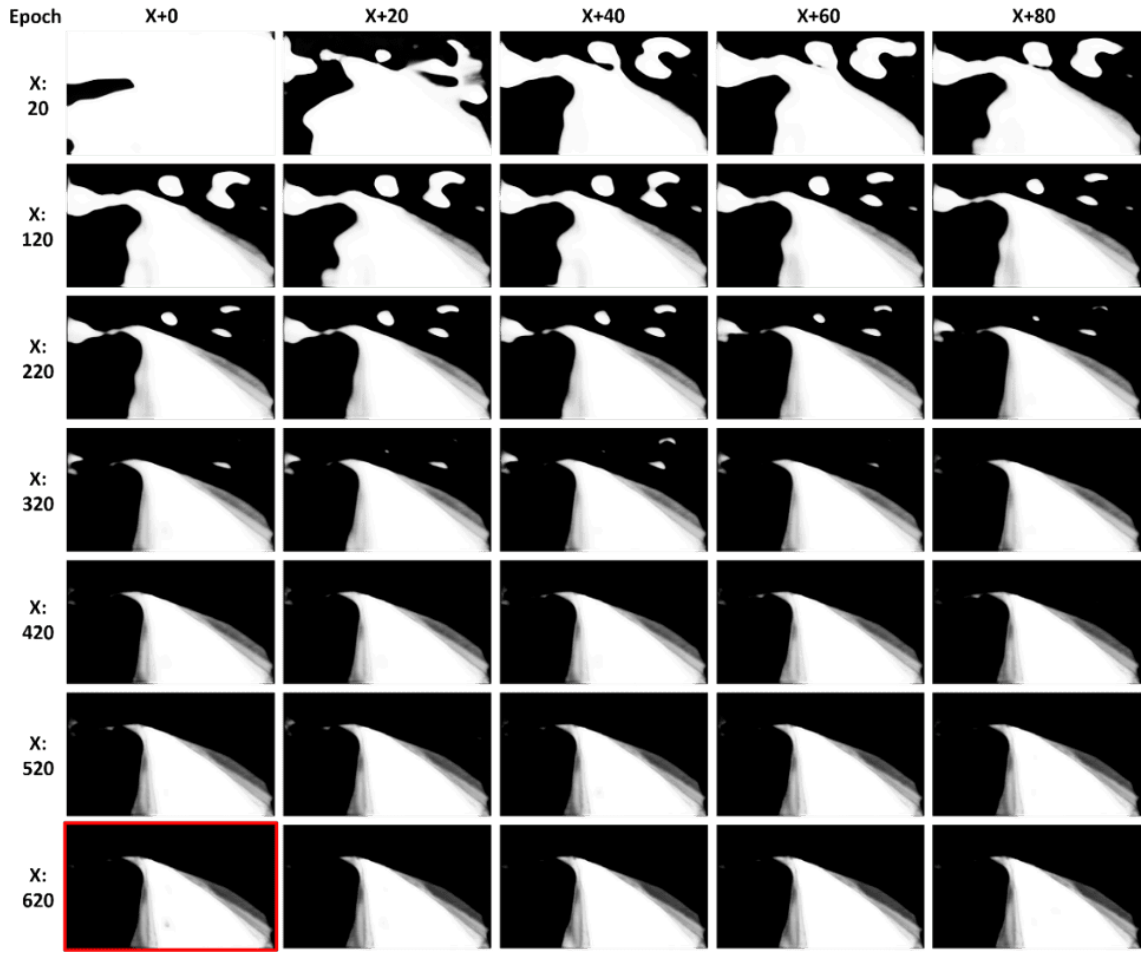

Figure 5. Illustration for the raw segmentation predictions of the task-specific U-Net model across optimizing epochs. Red bounding box indicates the result associated with the final optimized task-specific U-Net model.

doi.org/10.32388/EZWLSN

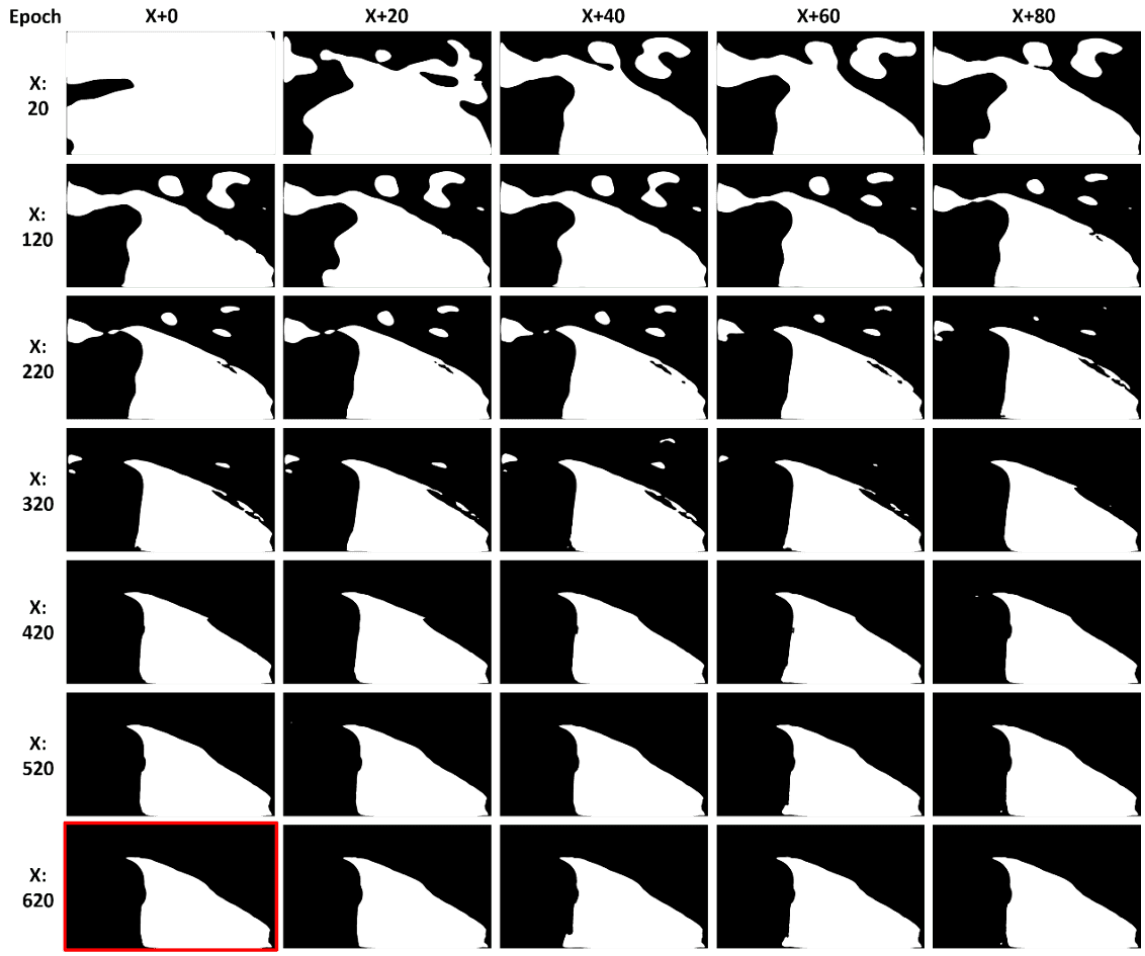

Figure 6. Illustration for binary segmentations by applying a 0.5 threshold to the raw predictions in Figure 5. Red bounding box indicates the result associated with the final optimized task-specific U-Net model.

doi.org/10.32388/EZWLSN

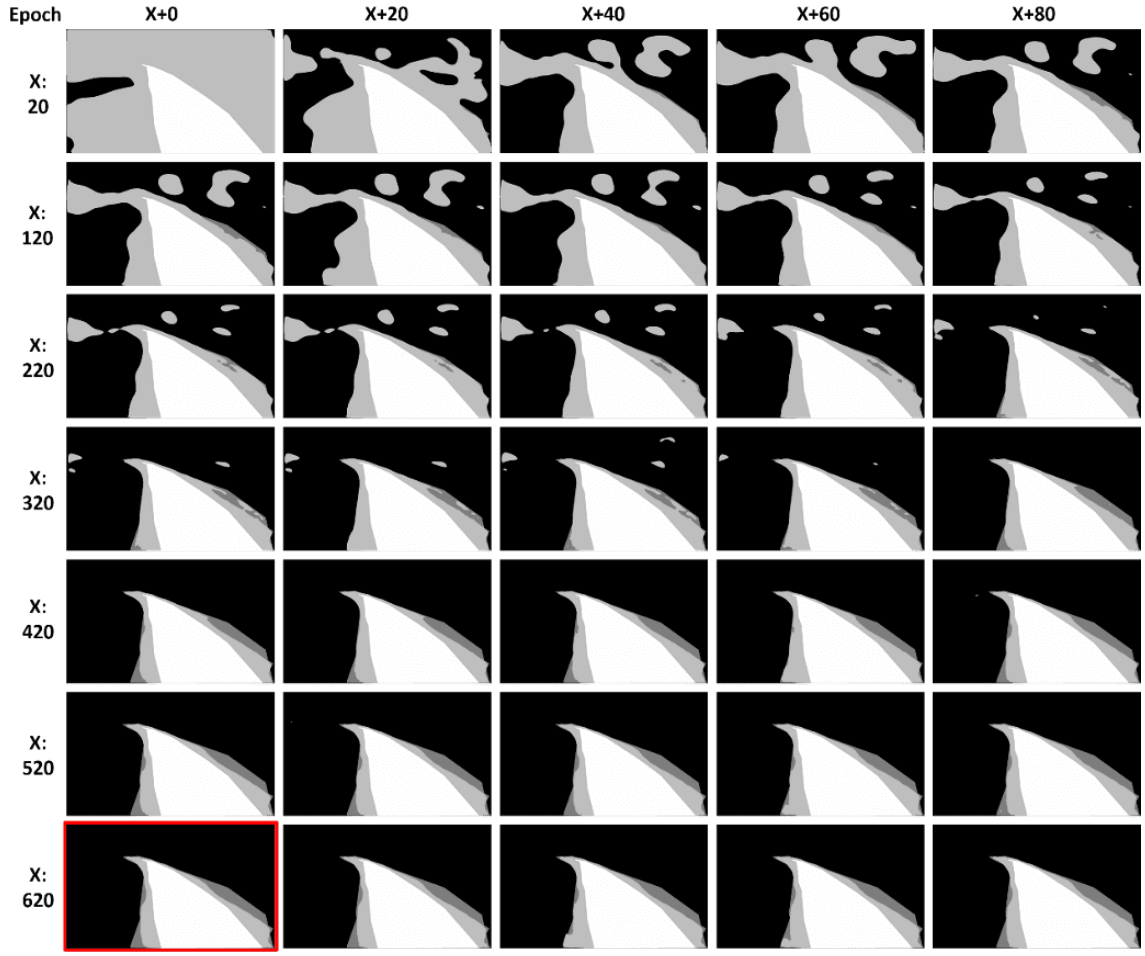

Figure 7. Illustrations for binary segmentations compared with the logical true target (LTT) defined in Formula (10). Red bounding box indicates the result associated with the final optimized task-specific U-Net model.

4.3.4. Qualitative comparison: Uncovered underlying true target vs. MIATTs

Beyond the quantitative evaluation in Section 4.3.2, qualitative comparisons were also conducted to further validate the effectiveness of the selected final optimized task-specific U-Net model. Particularly, the model's predicted latent true target is compared with the MIATTs corresponding to the input street image to offer direct visual evidence of its ability to uncover a reasonable and coherent representation of the underlying target. Fig. 8 illustrates the alignment between the predicted true target binarized from the distribution prediction for the latent true target and the logical true target (LTT) derived from the MIATTs, while Fig. 9 presents side-by-side comparisons between the predicted true target (binarized distribution prediction) and each individual inaccurate true target in the MIATTs set. Together, these visualizations are presented to show that the final model not only achieves strong numerical performance but also produces spatially plausible segmentations aligned with the logical constraints defined by the LTT.

doi.org/10.32388/EZWLSN

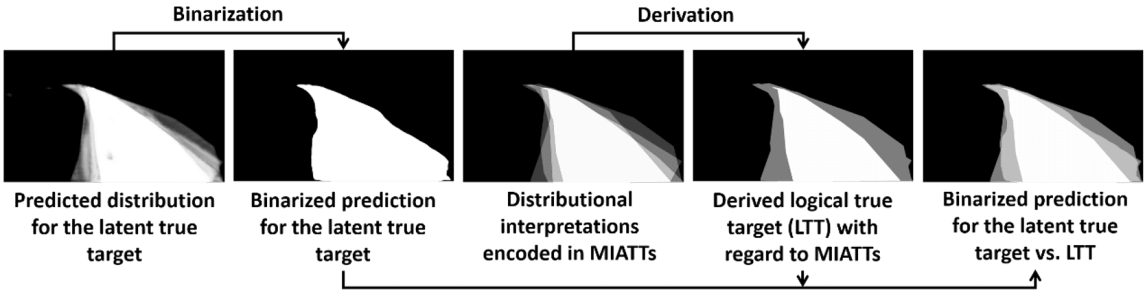

Figure 8. Visual comparison between the predicted true target binarized from the distribution prediction of the selected final optimized task-specific U-Net model for the latent true target and the logical true target (LTT) derived from the MIATTs. The binarized prediction (threshold = 0.5) closely aligns with the LTT, showing that the model effectively constrains its segmentation within the logical boundaries inferred from multiple inaccurate true targets.

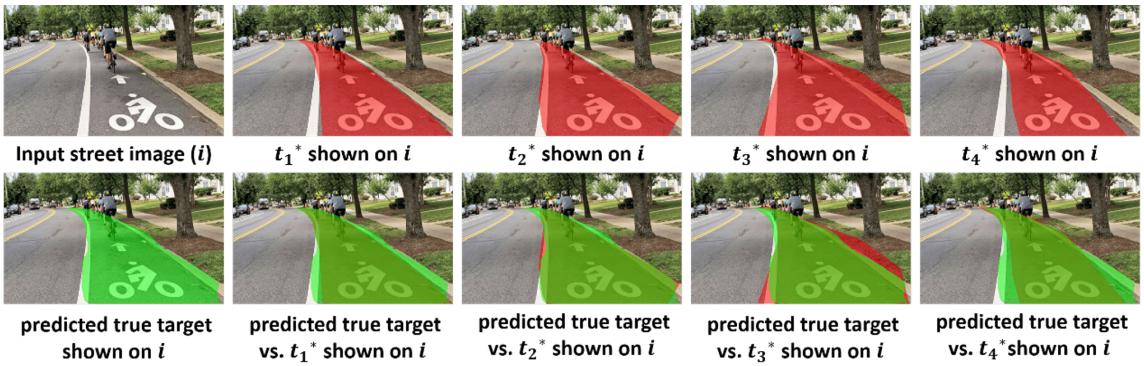

Figure 9. Visual comparison between the predicted latent true target of the selected final optimized task-specific U-Net model and each individual inaccurate true target within the MIATTs corresponding to the input street image. The comparisons reveal that the model's output integrates complementary and consistent information from the MIATTs set, while filtering out noise and biases present in individual inaccurate true targets. The input street image was photographed by Chris Roell and is used here with acknowledgment.

4.4. Analysis

The learning results summarized in Table 4 demonstrate a clear, monotonic, and stable improvement in model performance as optimizing progressed across epochs, reflecting effective convergence of the optimization process. At the early stage (epoch 20), the model exhibited low logical precision (LPrecision = 0.263) and modest logical intersection-over-union (LIoU = 0.263), accompanied by an extremely high count of logical false positives (LFP = 930,668), despite achieving perfect logical recall (LRecall = 1.000). This imbalance indicates that, although the model was highly sensitive to detecting potential true target pixels, it lacked sufficient discrimination, resulting in substantial over-segmentation. As training advanced, these shortcomings were progressively mitigated: LPrecision exceeded 0.80 by epoch 220, LIoU surpassed 0.90 by epoch 300, and from epoch 400 onward, LIoU consistently remained above 0.99. Logical error counts

doi.org/10.32388/EZWLSN

(LErrors) dropped precipitously during this phase, stabilizing in the low hundreds. The optimal LErrors was achieved at epoch 620, where LPrecision, LRecall, LF1, and LAccuracy all reached their maximum values (1.000 for most), and loss minimized to 0.173. Beyond this point, performance metrics plateaued with only minor oscillations in logical error counts, indicating that the model had effectively converged, and further training yielded negligible gains. This convergence pattern underscores the stability and reliability of the learned parameters for the task-specific U-Net model in reasonably predicting the underlying true target.

The trends depicted in Fig. 4 show the LIoU score has a consistent upward trajectory, rising from approximately 0.25 at epoch 20 to surpassing 0.999 after epoch 500. In parallel, the loss decreases steadily from about 0.70 to roughly 0.20, indicating progressive improvements in model fit. Total logical error counts (LErrors = LFP + LFN) remain above 1,300 during the early training stages but decline sharply after ~400 epochs, reaching a minimum of 53 at epoch 620. This point precisely satisfies the predefined stopping criteria (LIoU > 0.999, LErrors < 100), marking it as the optimal checkpoint. Selection of this epoch is further supported by the observed convergence pattern: performance metrics plateau beyond epoch 620, and logical error counts exhibit slight upward fluctuations thereafter, suggesting diminishing returns from continued training and a risk of overfitting. Thus, epoch 620 represents not only a quantitative optimum but also a strategically stable choice for balancing performance and generalization.

The visual trace provided in Fig. 5 shows how the predicted distribution of the underlying true target evolves during optimization. In early epochs, predictions appear coarse and contain substantial mis-segmented regions, reflecting high uncertainty and noise in the learned representation. As training progresses, the predicted masks gradually refine—boundaries become sharper, and spurious regions diminish—indicating the model's increasing ability to capture stable and discriminative features. By later epochs, the predictions converge to a consistent segmentation pattern, suggesting that the model has distilled a stable approximation of the latent true target's distribution from the MIATTs corresponding to the given input street image. This progression visually confirms the dynamic adjustment of the learned distribution toward a more accurate underlying true target representation contained in the MIATTs.

Visual traces depicted in Figs. 6 and 7 together demonstrate how binary segmentations, obtained by applying a 0.5 threshold to the raw predictions in Fig. 5, progressively align with the LTT defined in Formula (10) as training advances. In the early epochs, the binarized outputs exhibit large deviations from the LTT, with over-segmentation and inclusion of irrelevant regions. Over successive epochs, these binary masks become more spatially consistent, with extraneous predictions gradually removed and boundaries better aligned with the LTT structure. By later epochs, the segmentations are effectively constrained within a plausible region defined by the LTT, indicating that the optimization process systematically refines the model's predictions to adhere to the logical constraints of the task while retaining fidelity to the underlying true target distribution.

The convergence trends contained in Table 4 and Fig. 5 are mirrored in the visual progression shown in Figs. 6-7, where raw predictions and their binarized counterparts evolve from noisy, over-extended segmentations toward stable, well-localized outputs that closely align with the logical true target (LTT). Over successive epochs, the distribution of predicted target regions becomes increasingly constrained within reasonable LTT boundaries, confirming that the

optimization not only improved quantitative metrics but also refined the spatial distribution of the learned representation of the underlying true target.

The visual comparisons in Figs. 8 and 9 provide strong qualitative evidence supporting both the reasonableness and interpretability of the selected final optimized task-specific U-Net model's predictions. In Fig 8, the binarized prediction of the latent true target aligns closely with the logical true target (LTT) derived from the MIATTs, indicating that the model has effectively learned to constrain its segmentation within the logically consistent boundaries inferred from multiple noisy and biased supervision sources. This demonstrates that the optimization process not only improved numerical performance but also yielded a prediction that is visually coherent with the consensus structure implied by the MIATTs. Fig. 9 further elucidates the interpretability of the model's output by comparing it against each individual inaccurate true target. The predicted segmentation captures the shared structural patterns present across the MIATTs while suppressing idiosyncratic errors, biases, or omissions found in individual targets. Together, these comparisons confirm that the model's output is not only statistically optimal according to the evaluation metrics but also visually and conceptually consistent with the underlying true target, thereby reinforcing confidence in its practical reasonability.

5. Practical implications for future application prospects

In this section, from the results of the conducted experiments and corresponding analysis, we derive some practical implications for future application prospects. The experimental findings, despite being derived from a single-sample case study, reveal several key implications for the broader applicability of the EL-MIATTs framework in real-world settings.

First, the framework's ability to progressively refine predictions toward the latent true target across optimization epochs demonstrates its adaptability to tasks where the true target is uncertain. This capability is particularly relevant for domains such as medical imaging, environmental monitoring, and infrastructure assessment, where annotation quality varies and conventional supervised learning may be hindered.

Second, the LAF-based metrics introduced in the EL-MIATTs framework—such as logical IoU (LIoU), logical Precision, and logical Recall—provide a robust mechanism for evaluating models against a logically consistent true target rather than purely relying on noisy or potentially biased true targets. This shift can reduce overfitting to inaccurate labels and improve the interpretability of results, which is critical for high-stakes decision-making scenarios.

Third, the convergence behavior observed in the study suggests that the framework can be integrated into iterative model development pipelines, offering clear stopping criteria based on logical error thresholds and LIoU plateaus. Such criteria not only optimize training efficiency but also minimize risks of overfitting, making the framework more computationally practical for large-scale deployment.

Finally, the qualitative visual comparisons—illustrated in Figs. 8 and 9—highlight the framework's transparency and explainability. These visual diagnostics can serve as an effective communication tool for stakeholders without technical expertise, fostering trust in model outputs and facilitating cross-disciplinary adoption.

Collectively, these implications point to the EL-MIATTs framework as a promising approach for domains requiring both predictive accuracy and logical consistency, paving the way for its extension to larger datasets and more complex tasks.

6. Discussion / Limitations

In this study, the EL-MIATTs framework was applied to the bicycle lane segmentation task in street images, with experimental validation conducted on a single representative sample. While the scope of this preliminary experiment is intentionally narrow, the results demonstrate the feasibility and potential effectiveness of EL-MIATTs for addressing real-world challenges where true targets are uncertain. This case study represents an important milestone in bridging the gap between theoretical formulation and practical deployment, showing that the proposed EL-MIATTs framework can generate an optimized predictive model for logical, interpretable, and high-quality predictions even under minimal supervision.

However, the extremely limited sample size naturally constrains the generalizability, statistical robustness, and coverage of the findings. A broader-scale empirical evaluation—encompassing diverse geographic regions, environmental conditions, and input complexities—will be essential for fully validating the framework's adaptability and resilience. Despite these limitations, the current study provides a concrete proof-of-concept that transforms the theoretical underpinnings of EL-MIATTs into actionable, testable outcomes. This initial success lays a solid foundation for future research, which will focus on systematically expanding the dataset, refining model components, and conducting rigorous comparative analyses against established methods to strengthen both the theoretical credibility and the real-world applicability of the framework.

Statements and Declarations

Acknowledgements

The author, Yongquan Yang, especially thanks his Yang Family, Chengdu, China, for providing him with financial support and mental encouragement to do this research. The input street image utilized in this research is acknowledged with gratitude to its photographer, Chris Roell.

References

1. a, b, c, d, e, f *Yang Y (2024). "Logical Assessment Formula and Its Principles for Evaluations with Inaccurate Ground-Truth Labels." Knowl Inf Syst. doi:10.1007/s10115-023-02047-6.*
2. a, b, c, d, e, f *Yang Y (2024). "Undefinable True Target Learning." doi:10.32388/KBK3P8.*
3. a, b, c, d *Yang Y (2025). "Evaluation and Learning with Multiple Inaccurate True Targets." Qeios. doi:10.32388/UMHEFG.*
4. ^ *Yang Y (2023). "Discovering Scientific Paradigms for Artificial Intelligence Alignment." doi:10.13140/RG.2.2.15945.52320.*
5. a, b, c *Minaee S, Boykov YY, Porikli F, Plaza AJ, Kehtarnavaz N, Terzopoulos D (2021). "Image Segmentation Using Deep Learning: A Survey." IEEE Trans Pattern Anal Mach Intell. 1–1. doi:10.1109/TPAMI.2021.3059968.*

6. a, b Asgari Taghanaki S, Abhishek K, Cohen JP, Cohen-Adad J, Hamarneh G (2021). "Deep Semantic Segmentation of Natural and Medical Images: A Review." Artif Intell Rev. **54**:137–178. doi:10.1007/s10462-020-09854-1.
7. a, b, c, d Kurz F, Azimi SM, Sheu C-Y, d'Angelo P (2019). "Deep Learning Segmentation and 3D Reconstruction of Road Marking s Using Multiview Aerial Imagery." IJGI. **8**:47. doi:10.3390/ijgi8010047.
8. a, b, c Liu W, Qiao S, Gao K, He H, Chapman MA, Xu L, Li J (2024). "Advancements in Road Lane Mapping: Comparative Fine-T uning Analysis of Deep Learning-based Semantic Segmentation Methods Using Aerial Imagery." doi:10.48550/ARXIV.2410.0 5717.
9. a, b Zhang YJ (1996). "A Survey on Evaluation Methods for Image Segmentation." Pattern Recognition. **29**:1335–1346. doi:10.10 16/0031-3203(95)00169-7.
10. a, b Wang Z, Wang E, Zhu Y (2020). "Image Segmentation Evaluation: A Survey of Methods." Artif Intell Rev. **53**:5637–5674. do i:10.1007/s10462-020-09830-9.
11. ^ Adams R, Bischof L (1994). "Seeded Region Growing." IEEE Trans Pattern Anal Machine Intell. **16**:641–647. doi:10.1109/34.29 5913.
12. ^ Long J, Shelhamer E, Darrell T (2015). "Fully Convolutional Networks for Semantic Segmentation." In: Proceedings of the IE EE Conference on Computer Vision and Pattern Recognition: pp. 3431–3440.
13. a, b Ronneberger O, Fischer P, Brox T (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation." In: Navab N, Hornegger J, Wells WM, Frangi AF (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer International Publishing: pp. 234–241. doi:10.1007/978-3-319-24574-4_28.
14. ^ Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2018). "DeepLab: Semantic Image Segmentation with Deep Conv olutional Nets, Atrous Convolution, and Fully Connected CRFs." IEEE Trans Pattern Anal Mach Intell. **40**:834–848. doi:10.110 9/TPAMI.2017.2699184.
15. ^ Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S, Schiele B (2016). "The Cityscapes Data set for Semantic Urban Scene Understanding." In: Proceedings of the IEEE Conference on Computer Vision and Pattern Reco gnition: pp. 3213–3223.
16. ^ Sheng S, Formosa N, Hossain M, Quddus M (2024). "Advancements in Lane Marking Detection: An Extensive Evaluation of Current Methods and Future Research Direction." IEEE Trans Intell Veh. **9**:6462–6473. doi:10.1109/TIV.2024.3369733.
17. ^ Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y, others (2023). "Segment Anything." In: Proceedings of the IEEE/CVF International Conference on Computer Vision: pp. 4015–4026.
18. a, b Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE (2017). "A Survey of Deep Neural Network Architectures and Their Applicat ions." Neurocomputing. **234**:11–26. doi:10.1016/j.neucom.2016.12.038.
19. a, b Miikkulainen R, Liang J, Meyerson E, Rawal A, Fink D, Francon O, Raju B, Shahrzad H, Navruzyan A, Duffy N, Hodjat B (2 024). "Evolving Deep Neural Networks." In: Artificial Intelligence in the Age of Neural Networks and Brain Computing. Elsevi er: pp. 269–287. doi:10.1016/B978-0-323-96104-2.00002-6.
20. a, b Yang Y, Bu H (2024). "Validation of the Practicability of Logical Assessment Formula for Evaluations with Inaccurate Grou nd-Truth Labels: An Application Study on Tumour Segmentation for Breast Cancer." Comput Artif Intell. **2**:1443. doi:10.5940

0/cai.v2i2.1443.

21. ^Rahman MA, Wang Y (2016). "Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation." In: International Symposium on Visual Computing. Springer: pp. 234–244.
22. ^Azad R, Heidary M, Yilmaz K, Hüttemann M, Karimijafarbigloo S, Wu Y, Schmeink A, Merhof D (2023). "Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook." doi:10.48550/ARXIV.2312.05391.
23. ^Wang Q, Ma Y, Zhao K, Tian Y (2022). "A Comprehensive Survey of Loss Functions in Machine Learning." Ann Data Sci. **9**:187–212. doi:10.1007/s40745-020-00253-5.
24. a, b, c, dYang Y, Li F, Wei Y, Chen J, Chen N, Alobaidi MH, Bu H (2024). "One-Step Abductive Multi-Target Learning with Diverse Noisy Samples and Its Application to Tumour Segmentation for Breast Cancer." Expert Systems with Applications. **251**:123923. doi:10.1016/j.eswa.2024.123923.
25. a, b, c, dYang Y, Yang Y, Chen J, Zheng J, Zheng Z (2024). "Handling Noisy Labels Via One-Step Abductive Multi-Target Learning and Its Application to Helicobacter Pylori Segmentation." Multimed Tools Appl. doi:10.1007/s11042-023-17743-2.
26. a, b, c, dYang Y, Yang Y, Yuan Y, Zheng J, Zhongxi Z (2020). "Detecting Helicobacter Pylori in Whole Slide Images Via Weakly Supervised Multi-Task Learning." Multimed Tools Appl. **79**:26787–26815. doi:10.1007/s11042-020-09185-x.
27. ^Amari S (1993). "Backpropagation and Stochastic Gradient Descent Method." Neurocomputing. **5**:185–196. doi:10.1016/0925-2312(93)90006-O.
28. ^Ruder S (2016). "An Overview of Gradient Descent Optimization Algorithms." doi:10.48550/ARXIV.1609.04747.
29. ^Bian K, Priyadarshi R (2024). "Machine Learning Optimization Techniques: A Survey, Classification, Challenges, and Future Research Issues." Arch Computat Methods Eng. doi:10.1007/s11831-024-10110-w.
30. ^Mao A, Mohri M, Zhong Y (2023). "Cross-Entropy Loss Functions: Theoretical Analysis and Applications." In: Krause A, Brunskill E, Cho K, Engelhardt B, Sabato S, Scarlett J (Eds.), Proceedings of the 40th International Conference on Machine Learning. PMLR: pp. 23803–23828. https://proceedings.mlr.press/v202/mao23b.html.
31. ^Kline DM, Berardi VL (2005). "Revisiting Squared-Error and Cross-Entropy Functions for Training Neural Network Classifiers." Neural Comput & Applic. **14**:310–318. doi:10.1007/s00521-005-0467-y.
32. ^Milletari F, Navab N, Ahmadi S-A (2016). "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation." In: 2016 Fourth International Conference on 3D Vision (3DV). Ieee: pp. 565–571.
33. ^Zhao R, Qian B, Zhang X, Li Y, Wei R, Liu Y, Pan Y (2020). "Rethinking Dice Loss for Medical Image Segmentation." In: 2020 IEEE International Conference on Data Mining (ICDM). Sorrento, Italy: IEEE: pp. 851–860. doi:10.1109/ICDM50108.2020.00094.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.