TD-DIMB: Text-Driven Dense Inverted Mamba Bottlenecks for Interactive Medical Image Segmentation

Mohamed Lamine Allaoui and Mohand Saïd Allili

A Universal Training datasets and experimental setup

A.1 Complete Dataset Specifications for Universal Generalization

The evaluation of medical image segmentation models across diverse clinical domains presents a fundamental challenge in establishing true generalizability. While task-specific evaluation provides valuable insights under controlled conditions, clinical reality demands robust adaptation across unseen imaging modalities, anatomical variations, and institutional protocols. To address this challenge, we developed a comprehensive universal training framework leveraging heterogeneous medical imaging datasets to evaluate TD-DIMB's cross-domain generalization capacity without architectural retraining.

Our universal training corpus represents a strategically curated collection spanning the breadth of clinical practice. Rather than simply aggregating available datasets, we selected collections that collectively represent major imaging modalities, anatomical systems, and pathological conditions encountered in modern radiology. This principled approach ensures our generalization evaluation reflects realistic clinical deployment scenarios while maintaining semantic flexibility essential for prompt-driven segmentation.

The foundation of our universal evaluation rests on fourteen distinct medical imaging collections, encompassing over 15,761 individual studies across three primary modalities and providing unprecedented scope for cross-domain learning evaluation. Table 1 presents complete specifications of our

Table 1: Universal Training Dataset Specifications

Dataset	Clinical Description	Studies	Modality	Primary Anatomical Focus
AbdomenCT-1K	Multi-organ abdominal segmentation with comprehensive anatomical coverage including hepatic, renal, and splenic structures	361	CT	Liver, kidneys, spleen, pancreas, gallbladder
AMOS	Cross-modal abdominal organ segmentation enabling validation of modality-invariant feature learning	240	CT, MRI	Abdominal organs with cross-modal consistency
BraTS	Comprehensive brain tumor segmentation including multiple tumor sub-regions and peritumoral structures	6,096	Multi-modal MRI	Gliomas, enhancing regions, tumor core, edema
CHAOS	Cross-modal abdominal organ challenge emphasizing consistency across imaging protocols	40	CT, T2-MRI	Liver, kidneys, spleen with modal alignment
KiTS 2023	Kidney and renal tumor segmentation with detailed pathological annotations	489	CT	Kidneys, renal tumors, cystic lesions
LiTS	Liver tumor segmentation focusing on precise boundary delineation in oncological imaging	131	CT	Liver parenchyma, hepatic lesions
LUNA	Pulmonary nodule analysis with comprehensive lung parenchyma coverage	888	CT	Lung tissue, pulmonary nodules
MSD Collection	Medical Segmentation Decathlon spanning ten distinct anatomical segmentation tasks	3,225	CT, Multi-modal MRI	Heart, hippocampus, liver, lung, pancreas, prostate, colon, vessels, spleen, brain tumors
PROMISE12	Prostate segmentation using high-resolution T2-weighted magnetic resonance imaging	37	T2-MRI	Prostate gland, peripheral zone
ABUS	Automated breast ultrasound with three-dimensional tumor characterization	200	3D Ultrasound	Breast lesions, fibrocystic changes
BUSI	Breast ultrasound imaging with comprehensive benign and malignant lesion representation	1312	2D Ultrasound	Breast masses, tumor boundaries
Total Segmentator V2	Anatomical atlas segmentation encompassing 117 distinct anatomical structures	1,228	CT	Comprehensive full-body anatomical mapping
CAMUS	Cardiac motion analysis using multi-view echocardiographic imaging	500	Ultrasound	Left ventricle, left atrium, myocardium
autoPET 2022	Whole-body lesion segmentation in oncological positron emission tomography	1,014	PET/CT	Metabolically active lesions, tumor burden assessment

universal training corpus, highlighting the diversity of anatomical targets and imaging protocols represented.

This heterogeneous composition serves multiple scientific purposes beyond simple modality variation. It enables evaluation of TD-DIMB's ability to learn modality-invariant anatomical representations that generalize across imaging physics, from high-resolution CT cross-sectional anatomy to MRI soft tissue characterization, ultrasound real-time acoustic imaging, and PET functional metabolic assessment. The diverse anatomical coverage tests the model's capacity to transfer learned features between different organ systems and pathological conditions, while variation in dataset sizes and annotation quality reflects realistic clinical data collection constraints.

The anatomical distribution within our universal corpus reflects clinical imaging prevalence while ensuring adequate representation across major organ systems. Neurological studies comprise the largest component at 40.6% of the corpus, primarily driven by the comprehensive BraTS brain tumor dataset, encompassing both structural and pathological brain imaging. Multi-organ anatomical atlas studies represent 24.4%, providing comprehensive full-body coverage through TotalSegmentator and MSD collections. Thoracic studies account for 10.7%, including both pulmonary and cardiac applications. Breast imaging represents 9.6% through ultrasound modalities, while abdominal imaging comprises 8.0% covering hepatic, renal, and splenic structures. Oncological lesion studies contribute 6.4% through whole-body PET/CT imaging. This distribution reflects the comprehensive scope of our universal training while providing sufficient diversity for robust cross-domain evaluation.

A.2 Universal Data Preparation and Text Prompt Generation

The preparation of medical imaging data for universal training presents unique challenges extending beyond conventional computer vision preprocessing. Medical images exhibit profound variations in intensity ranges, spatial resolutions, and acquisition parameters reflecting underlying imaging physics, while clinical interpretation relies on domain-specific knowledge that must be preserved throughout preprocessing. Our approach addresses these challenges through a comprehensive framework that normalizes technical variations while preserving clinically-relevant information.

A.2.1 Strategic Slice Extraction and Multi-Label Task Construction

Given TD-DIMB's two-dimensional architecture, we extract representative slices from three-dimensional volumes using a dual-strategy approach balancing anatomical coverage with label representation. We extract two distinct slice types: mid-slices and max-slices, each serving specific training purposes.

Mid-slice extraction systematically captures the middle slice along each anatomical axis (axial, sagittal, coronal), resulting in representative 256×256 cross-sections. This strategy avoids bias toward pathological regions, ensuring TD-DIMB learns robust anatomical representations that generalize across spatial locations, critical for clinical deployment where structure locations are unknown *a priori*.

Max-slice extraction supplements training data by identifying slices containing maximum voxel counts for each anatomical label l. This addresses label sparsity common in medical imaging, where structures may occupy minimal volume fractions, providing concentrated anatomical examples while maintaining context.

A.2.2 Combinatorial Prompt Generation and Multi-Class Task Decomposition

A critical innovation lies in systematically generating anatomically-grounded text prompts reflecting complex multi-class medical annotations. Rather than treating multi-class segmentation as single tasks, we decompose annotated slices into multiple prompt-driven tasks capturing individual structures and clinically-relevant combinations.

For datasets with multiple anatomical labels per slice, we implement combinatorial task generation creating individual and composite targets. The CAMUS cardiac dataset exemplifies this approach: containing left ventricular epicardium, endocardium, and left atrium, we generate binary tasks reflecting clinical reasoning patterns rather than single three-class segmentation.

Individual structures receive dedicated binary masks with corresponding prompts ("Segment the left ventricular epicardium," "Delineate the left atrial boundary"), enabling precise anatomical discrimination with clinical terminology consistency. Combinatorial tasks reflect meaningful anatomical relationships through composite prompts ("Segment the epicardium and

endocardium"), with binary masks encompassing specified region unions.

This approach serves multiple purposes: significantly expanding training corpus size, enabling flexible semantic associations across anatomical specificity levels, and mirroring clinical workflows where radiologists focus on individual or combined structures based on clinical questions.

Mathematically, for slices containing n anatomical labels $\{l_1, l_2, \ldots, l_n\}$, we generate $2^n - 1$ distinct tasks corresponding to all non-empty label subsets. Each task t_S for subset S receives binary mask $M_S = \bigcup_{l_i \in S} M_{l_i}$ and corresponding textual prompt P_S describing the anatomical union in clinically-appropriate language.

B Reinforced Gaussian Dice Loss (RGDL) Extended Analysis

B.1 Mathematical Formulation for Binary Segmentation

B.1.1 Notation and Preliminaries

We establish the following mathematical notation for the segmentation task:

- $X \in \mathbb{R}^{B \times C \times H \times W}$: Batch of input images, where B denotes the batch size, C represents the number of channels, and H, W correspond to the height and width dimensions
- $Y \in \{0, 1\}^{B \times 1 \times H \times W}$: Original binary ground truth segmentation masks
- $\tilde{Y} \in [0, 1]^{B \times 1 \times H \times W}$: Gaussian-smoothed ground truth masks
- $\hat{Y} \in [0, 1]^{B \times 1 \times H \times W}$: Predicted probability maps
- $f_\theta : X \to \hat{Y}$: Segmentation model parameterized by θ

B.1.2 Gaussian Smoothing of Ground Truth

To account for annotation uncertainty, we apply Gaussian smoothing to the binary ground truth masks:

$$\tilde{Y} = Y * G_\sigma \tag{1}$$

where $*$ represents the convolution operation and G_σ is a Gaussian kernel with standard deviation σ:

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

For a specific pixel location (i, j), the smoothed ground truth is computed as:

$$\tilde{Y}(i,j) = \sum_{m=-k}^{k} \sum_{n=-k}^{k} Y(i+m, j+n) \cdot G_\sigma(m,n) \tag{3}$$

with k representing the kernel radius (typically $k = 3\sigma$).

This smoothing operation converts hard binary masks to soft probability masks, with values ranging from 0 to 1, representing the confidence or probability that a pixel belongs to the foreground. Pixels at object boundaries will have intermediate values, reflecting the uncertainty in the exact boundary location, resulting in a dilated mask.Figure 1 illustrates the effect of different sigma values on mask dilation, demonstrating how Gaussian smoothing progressively expands the segmentation boundary as sigma increases.

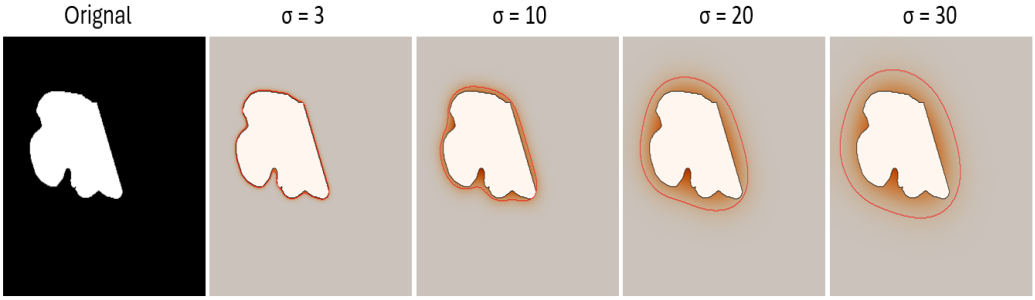

Figure 1: Visualization of Gaussian-based mask dilation at 256×256 resolution. From left to right: original binary mask, followed by dilated masks with increasing sigma values ($\sigma = 3, 10, 20, 30$). Orange regions show the probability field generated by Gaussian smoothing, with lighter to darker orange indicating lower to higher probability values. Red contours mark the boundary of the dilated mask at 0.5 threshold. Note how larger sigma values produce more extensive boundary expansion while preserving the overall shape characteristics of the original segmentation.

B.1.3 Reinforced Gaussian Dice Loss Formulation

The Reinforced Gaussian Dice Loss is formulated as a weighted combination of two complementary components:

$$\mathcal{L}_{\text{RGDL}}(\tilde{Y}, Y, \hat{Y}) = \lambda_1 \cdot \mathcal{L}_{\text{Dice}}(\tilde{Y}, Y, \hat{Y}) + \lambda_2 \cdot \mathcal{L}_{\text{RL}}(Y, \hat{Y}) \tag{4}$$

where $\lambda_1, \lambda_2 \in [0, 1]$ are weighting coefficients with the constraint $\lambda_1 + \lambda_2 = 1$, enabling adaptive balancing between global and local optimization objectives.

Modified Dice Loss Component. We modify the Dice loss term to use the smoothed ground truth only in the numerator while retaining the original binary mask in the denominator:

$$\mathcal{L}_{\text{Dice}}(\tilde{Y}, Y, \hat{Y}) \;=\; 1 \;-\; \frac{2\sum_{i=1}^{N} \tilde{y}_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i \;+\; \sum_{i=1}^{N} \hat{y}_i \;+\; \epsilon}, \tag{5}$$

where $\tilde{y}_i$, y_i, and $\hat{y}_i$ represent individual elements of $\tilde{Y}$, Y, and $\hat{Y}$ respectively, ϵ constitutes a small smoothing constant to ensure numerical stability by preventing division by zero, and $N = B \times H \times W$ denotes the total number of pixels across the batch.

This formulation offers a significant advantage: it encourages the model to extend predicted regions slightly beyond the hard boundaries (improving recall) while still maintaining volumetric consistency with the original mask. By using the original binary mask in the denominator, we ensure that the total volume of the prediction remains comparable to the ground truth, preventing excessive over-segmentation that might occur if the smoothed mask were used in both numerator and denominator.

Learnable Reinforcement Parameters. To enhance the adaptability of our loss function, we replace the fixed hyperparameters with learnable parameters. We define:

$$\alpha = \alpha_{\max} \cdot \sigma(\theta_\alpha) \tag{6}$$

$$\beta = \beta_{\max} \cdot \sigma(\theta_\beta) \tag{7}$$

$$\gamma = \gamma_{\max} \cdot \sigma(\theta_\gamma) \tag{8}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function, ensuring that the parameters remain within a reasonable range. The terms $\alpha_{\max}$, $\beta_{\max}$, and $\gamma_{\max}$ are scaling constants that define the maximum values these parameters can reach. The parameters θ_α, θ_β, and θ_γ are trainable network parameters that are learned through backpropagation.

Reinforcement-Learning Component. To systematically incorporate pixel-wise incentives and penalties with learnable parameters, we define a reward function that differentiates between various prediction outcomes:

$$r(y_i, \hat{y}_i) = \begin{cases} +\alpha \cdot y_i \hat{y}_i & \text{(true positive contribution)}, \\ -\beta \cdot (1 - y_i)\hat{y}_i & \text{(false positive contribution)}, \\ -\gamma \cdot y_i(1 - \hat{y}_i) & \text{(false negative contribution)}, \end{cases} \tag{9}$$

We use the original binary ground truth values y_i to maintain clear and targeted reinforcement signals that precisely distinguish between different prediction outcomes.

Based on this reward structure, we define the reinforcement learning loss term by aggregating the rewards/penalties and normalizing by the total number of positive (foreground) pixels in the original ground truth:

$$\mathcal{L}_{\text{RL}}(Y, \hat{Y}) \;=\; -\frac{1}{\sum_{i=1}^{N} y_i + \epsilon}\Big(\alpha \sum_{i=1}^{N} y_i \hat{y}_i \;-\; \beta \sum_{i=1}^{N} (1 - y_i)\, \hat{y}_i \;-\; \gamma \sum_{i=1}^{N} y_i (1 - \hat{y}_i)\Big). \tag{10}$$

The negative sign in this formulation translates the reinforcement learning paradigm of "maximizing reward" into the optimization framework of "minimizing loss."

Combined Reinforced Gaussian Dice Loss. Integrating the modified Dice and reinforcement learning components, the complete Reinforced Gaussian Dice Loss is expressed as:

$$\mathcal{L}_{\text{RGDL}}(\tilde{Y}, Y, \hat{Y}) = \lambda_1 \Big[1 - \frac{2 \sum_{i=1}^{N} \tilde{y}_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \epsilon}\Big] \tag{11}$$

$$- \lambda_2 \frac{\alpha \sum_{i=1}^{N} y_i \hat{y}_i \;-\; \beta \sum_{i=1}^{N} (1 - y_i)\hat{y}_i \;-\; \gamma \sum_{i=1}^{N} y_i (1 - \hat{y}_i)}{\sum_{i=1}^{N} y_i + \epsilon} \tag{12}$$

where α, β, and γ are the learnable parameters as defined above. This formulation establishes a sophisticated balance between global structural overlap assessment (via the Dice component) and localized pixel-wise supervisory signals (via the reinforcement learning component), while also accounting for annotation uncertainty through Gaussian smoothing.

B.1.4 Differentiability and Gradients

The Reinforced Gaussian Dice Loss maintains full differentiability with respect to the network parameters θ, as each component exhibits smooth dependence on the predicted probabilities $\hat{y}_i$. For a specific pixel's prediction $\hat{y}_j$, the gradient is computed as:

$$\frac{\partial \mathcal{L}_{\text{RGDL}}}{\partial \hat{y}_j} = \lambda_1 \frac{\partial \mathcal{L}_{\text{Dice}}}{\partial \hat{y}_j} + \lambda_2 \frac{\partial \mathcal{L}_{\text{RL}}}{\partial \hat{y}_j}. \tag{13}$$

Gradient of the Modified Dice Component. For the modified Dice term $\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum \tilde{y}_i \hat{y}_i + \epsilon}{\sum y_i + \sum \hat{y}_i + \epsilon}$, application of the quotient rule yields:

$$\frac{\partial \mathcal{L}_{\text{Dice}}}{\partial \hat{y}_j} = -\frac{\partial}{\partial \hat{y}_j}\Big[\frac{2\sum_{i=1}^{N} \tilde{y}_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \epsilon}\Big] \tag{14}$$

$$= -\frac{2\,\tilde{y}_j\big(\sum y_i + \sum \hat{y}_i + \epsilon\big) \;-\; \big(2\sum_{i=1}^{N} \tilde{y}_i \hat{y}_i + \epsilon\big)}{\big(\sum y_i + \sum \hat{y}_i + \epsilon\big)^2}. \tag{15}$$

This gradient formulation reveals a key advantage: pixels near object boundaries where $\tilde{y}_j > 0$ receive a stronger positive gradient signal than they would with the conventional Dice loss. This specifically encourages the model to expand its predictions slightly beyond the hard boundary defined in the original binary mask, improving recall without compromising overall volumetric consistency.

Gradient of the RL Component. For the reinforcement learning term with original binary ground truth

$$\mathcal{L}_{\text{RL}} = -\frac{1}{\sum y_i + \epsilon}\Big[\alpha \sum y_i \hat{y}_i - \beta \sum (1 - y_i)\hat{y}_i - \gamma \sum y_i (1 - \hat{y}_i)\Big],$$

we conduct a case-based analysis depending on whether the pixel belongs to the foreground or background:

Case 1: $y_j = 1$ **(Foreground Pixel).** For foreground pixels, only the terms $\alpha \sum y_i \hat{y}_i$ and $-\gamma \sum y_i (1 - \hat{y}_i)$ contribute to the gradient calculation for pixel j. The derivative is computed as:

$$\frac{\partial \mathcal{L}_{\text{RL}}}{\partial \hat{y}_j} = -\frac{1}{\sum y_i + \epsilon}\Big[\alpha \frac{\partial}{\partial \hat{y}_j}(y_j \hat{y}_j) + \big(-\gamma\big) \frac{\partial}{\partial \hat{y}_j}\big(y_j (1 - \hat{y}_j)\big)\Big]. \tag{16}$$

Given that $y_j = 1$, we have $\frac{\partial}{\partial \hat{y}_j}(y_j \hat{y}_j) = 1$ and $\frac{\partial}{\partial \hat{y}_j}(y_j(1 - \hat{y}_j)) = -1$. Consequently, the expression within the brackets simplifies to $\alpha + \gamma$, resulting in:

$$\frac{\partial \mathcal{L}_{\text{RL}}}{\partial \hat{y}_j} = -\frac{\alpha + \gamma}{\sum_{i=1}^{N} y_i + \epsilon} \quad (\text{for } y_j = 1).$$

This formulation reveals an important insight: both increasing true positives (rewarded by α) and decreasing false negatives (penalized by γ) contribute additively to the gradient signal, enhancing the model's ability to identify foreground structures.

Case 2: $y_j = 0$ (Background Pixel). For background pixels, only the term $-\beta \sum (1 - y_i)\hat{y}_i$ contributes to the gradient calculation for pixel j:

$$\frac{\partial \mathcal{L}_{\text{RL}}}{\partial \hat{y}_j} = -\frac{1}{\sum y_i + \epsilon}\Big[-\beta \frac{\partial}{\partial \hat{y}_j}\big((1 - y_j)\hat{y}_j\big)\Big].$$

Since $y_j = 0$ implies $(1 - y_j) = 1$, the derivative $\frac{\partial}{\partial \hat{y}_j}(\hat{y}_j) = 1$, yielding:

$$\frac{\partial \mathcal{L}_{\text{RL}}}{\partial \hat{y}_j} = +\frac{\beta}{\sum_{i=1}^{N} y_i + \epsilon}.$$

This positive gradient indicates that increasing the prediction value for background pixels is appropriately penalized by factor β, discouraging false positive predictions.

Gradients for Learnable Parameters. The loss function also produces gradients with respect to the learnable reinforcement parameters:

$$\frac{\partial \mathcal{L}_{\text{RGDL}}}{\partial \theta_\alpha} = \frac{\partial \mathcal{L}_{\text{RGDL}}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \theta_\alpha} \tag{17}$$

$$\frac{\partial \mathcal{L}_{\text{RGDL}}}{\partial \theta_\beta} = \frac{\partial \mathcal{L}_{\text{RGDL}}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \theta_\beta} \tag{18}$$

$$\frac{\partial \mathcal{L}_{\text{RGDL}}}{\partial \theta_\gamma} = \frac{\partial \mathcal{L}_{\text{RGDL}}}{\partial \gamma} \cdot \frac{\partial \gamma}{\partial \theta_\gamma} \tag{19}$$

With the sigmoid parameterization:

$$\frac{\partial \alpha}{\partial \theta_\alpha} = \alpha_{\max} \cdot \sigma(\theta_\alpha) \cdot (1 - \sigma(\theta_\alpha)) \tag{20}$$

$$\frac{\partial \beta}{\partial \theta_\beta} = \beta_{\max} \cdot \sigma(\theta_\beta) \cdot (1 - \sigma(\theta_\beta)) \tag{21}$$

$$\frac{\partial \gamma}{\partial \theta_\gamma} = \gamma_{\max} \cdot \sigma(\theta_\gamma) \cdot (1 - \sigma(\theta_\gamma)) \tag{22}$$

These gradients enable the reinforcement parameters to be learned through backpropagation along with the network weights.

B.2 Hyperparameter Configuration

The Reinforced Gaussian Dice Loss requires configuration of several hyperparameters to optimize performance for specific segmentation tasks:

- $\lambda_1, \lambda_2 \in [0, 1]$: These coefficients control the relative weighting between the Dice and reinforcement learning components, typically constrained such that $\lambda_1 + \lambda_2 = 1$ to maintain a balanced optimization objective.
- σ: This parameter controls the standard deviation of the Gaussian kernel used for smoothing the ground truth masks. Larger values of σ create more uncertainty around boundaries, which can help in cases where annotations are particularly variable. Typical values range from 0.5 to 2.0 pixels.
- $\alpha_{\max}$: This parameter determines the maximum reward magnitude for true positive predictions, generally set within the range of 1.0-2.0.
- $\beta_{\max}$: This coefficient establishes the maximum penalty magnitude for false positive predictions, typically configured within the range of 0.5-1.5.
- $\gamma_{\max}$: This parameter defines the maximum penalty magnitude for false negative predictions, usually set within the range of 0.5-1.5.
- ϵ: This smoothing constant, typically on the order of 10^{-5}, ensures numerical stability during optimization, particularly for small structures or in early training stages.

Parameter Initialization. The learnable parameters θ_α, θ_β, and θ_γ are initialized such that the initial values of α, β, and γ align with domain knowledge about the relative importance of different types of errors:

$$\theta_\alpha = \sigma^{-1}\left(\frac{\alpha_{\text{init}}}{\alpha_{\max}}\right) \tag{23}$$

$$\theta_\beta = \sigma^{-1}\left(\frac{\beta_{\text{init}}}{\beta_{\max}}\right) \tag{24}$$

$$\theta_\gamma = \sigma^{-1}\left(\frac{\gamma_{\text{init}}}{\gamma_{\max}}\right) \tag{25}$$

where $\sigma^{-1}(p) = \log\left(\frac{p}{1-p}\right)$ is the inverse sigmoid function, and α_{init}, β_{init}, and γ_{init} are the desired initial values for these parameters.

We recommend the following initialization strategy for medical image segmentation tasks:

$$\alpha_{\max} = 2.0, \qquad \alpha_{\text{init}} = 1.0 \tag{26}$$

$$\beta_{\max} = 1.5, \qquad \beta_{\text{init}} = 0.7 \tag{27}$$

$$\gamma_{\max} = 1.5, \qquad \gamma_{\text{init}} = 1.3 \tag{28}$$

This strategic initialization offers several key advantages:

- Setting $\gamma_{\text{init}} > \beta_{\text{init}}$ (1.3 ¿ 0.7) initially prioritizes recall over precision, which aligns with the clinical preference in medical image segmentation where false negatives (missing part of a lesion) are typically more problematic than false positives (including some surrounding tissue).
- The moderate value of $\alpha_{\text{init}} = 1.0$ provides a balanced starting point for true positive rewards, allowing the model to appropriately focus on correct classifications from the beginning of training.
- The chosen maximum values ($\alpha_{\max} = 2.0$, $\beta_{\max} = 1.5$, $\gamma_{\max} = 1.5$) provide sufficient room for the parameters to adapt during training while preventing them from reaching extreme values that might destabilize the optimization process.

For multi-class segmentation tasks, class-specific initializations can be employed. For small or clinically critical structures, higher initial values of γ (e.g., $\gamma_{\text{init}} = 1.5$) can be used to more strongly penalize false negatives from the beginning of training.

B.3 Extension to Multi-class Segmentation

For segmentation tasks involving multiple classes, the Reinforced Gaussian Dice Loss can be naturally extended to accommodate K-class problems. Let

Y_k, $\tilde{Y}_k$ and $\hat{Y}_k$ represent the original binary ground-truth mask, smoothed ground-truth mask, and predicted probability map for class k, respectively. The multi-class formulation is defined as:

$$\mathcal{L}_{\text{MC-RGDL}}(\tilde{Y}, Y, \hat{Y}) = \sum_{k=0}^{K-1} w_k \Big[\lambda_1 \, \mathcal{L}_{\text{Dice}}(\tilde{Y}_k, Y_k, \hat{Y}_k) + \lambda_2 \, \mathcal{L}_{\text{RL}}(Y_k, \hat{Y}_k) \Big], \quad (29)$$

where $w_k \geq 0$ represents the weight for class k with the constraint $\sum_{k=0}^{K-1} w_k = 1$.

This formulation enables class-specific error prioritization by allowing each class to have its own set of learnable reinforcement parameters $(\alpha_k, \beta_k, \gamma_k)$, which is especially valuable in medical imaging contexts where different anatomical structures or pathological regions may have varying clinical significance.