

HadBalance: A Plug-and-Play Unified Global Geometric Prior Framework for Generalizable Biomedical Segmentation

Zhuangzhi Gao^{1,2}, Feixiang Zhou¹, He Zhao¹, Wenhan Chen⁴, Ruiyu Luo¹, Xin Wang³, Hongyi Qin⁵, Zhongli Wu^{1,8}, Yanda Meng⁶, Yitian Zhao⁷, Alena Shantsila⁸, Gregory Y. H. Lip⁸, Eduard Shantsila², and Yalin Zheng^{1*}

¹ Department of Eye and Vision Science, University of Liverpool, Liverpool, UK
yalin.zheng@liverpool.ac.uk

² Department of Primary Care and Mental Health, University of Liverpool, Liverpool, UK

³ School of Psychological and Cognitive Sciences, Peking University, Beijing, China

⁴ Computer Vision Research Group, University of Amsterdam, Netherlands

⁵ Institute of Life Course & Medical Sciences, University of Liverpool, Liverpool, UK

⁶ Bioengineering Program, Biomedical Sciences Division (BioMed), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

⁷ Ningbo Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences, Cixi, China

⁸ Liverpool Centre for Cardiovascular Science, University of Liverpool, Liverpool, UK

Abstract. Precise biomedical image segmentation is crucial for clinical diagnosis. Geometric cues (e.g., boundary, shape, and topology) can improve structural consistency, yet most are task-specific and lack a unified geometric foundation that generalizes across organs and modalities. We are motivated by the observation that many medical segmentation targets exhibit globally near-convex shapes. A convex region is one in which any two interior points can be connected by a line segment entirely contained within the region. In practice, medical targets may exhibit small local concavities or boundary irregularities; we refer to such globally convex-like shapes as near-convex. Motivated by this, we derive *Hadwiger Shape Priors* from Hadwiger’s theorem as an interpretable global regularizer using three 2D measures: area A , perimeter P , and Euler characteristic χ , enabling transfer across organs and modalities. However, because medical datasets are shape-heterogeneous, enforcing near-convex priors uniformly can over-regularize non-convex anatomy (with significant concavities), washing out concavities and fine details and degrading segmentation accuracy. To address this challenge, we propose *Conflict-Aware Objective Balancing (CAOB)*, which integrates shape priors with segmentation in a gradient-aware manner. For each prior, CAO removes only the gradient component that conflicts with segmentation while preserving the remaining aligned component, and adaptively regulates objective influences to prevent prior dominance. This enables stable use of shape priors on shape-heterogeneous data

* Corresponding author.

without erasing genuine concavities or fine structural details. We call this plug-and-play framework ***HadBalance***. Our code is available at: <https://github.com/NatsuGao7/HadBalance>.

Keywords: Segmentation · Polyps · Carotid · Optic Disc

1 Introduction

Accurate biomedical image segmentation is a cornerstone of clinical diagnosis and has attracted research efforts in recent years, leading to deep learning models tailored to specific organs and imaging modalities that achieve state-of-the-art performance [8, 10, 17, 25, 26, 30]. Incorporating geometric cues (e.g., boundary [29], shape [28], and topology [9]) is a common way to improve structural consistency; however, existing geometric priors are largely task-specific (e.g., topology for vessel connectivity [13] and boundary terms for polyp delineation [29]) and ***lack a unified foundation that generalizes across most organs and modalities and supports transferable geometric constraints***.

We are motivated by the observation that many medical segmentation targets can be approximated as globally near-convex shapes [15, 16, 22]. A convex region is one in which any two interior points can be connected by a line segment entirely contained within the region [4]. As shown in Fig. 1(A), common anatomical structures—such as polyp regions [3], the carotid lumen in ultrasound [18], and the optic disc [1]—are largely convex in their overall shape, despite minor local concavities; we refer to such globally convex-like shapes as near-convex. Motivated by this observation, we draw on Hadwiger’s theorem [11], which provides a complete and minimal basis for shape description via three intrinsic volumes (in 2D: area A , perimeter P , and Euler characteristic χ ; see Section 2). Accordingly, we propose *Hadwiger Shape Priors* that impose complementary global constraints on the prediction via A , P , and χ .

However, ***due to data heterogeneity, near-convex and non-convex anatomies with significant concavities often coexist within the same dataset*** (Fig. 1(A)). When applied uniformly, Hadwiger Shape Priors may over-regularize non-convex anatomies, suppressing genuine concavities and fine boundary details. A common remedy is to gate priors using heuristic criteria or auxiliary networks; however, reliably distinguishing near-convex from non-convex cases in a unified and generalizable manner is inherently challenging, often introducing additional hyperparameters and unstable decisions on ambiguous samples [28]. A more principled solution is to address the issue at the optimization level: when shape priors conflict with the segmentation objective, analyzing gradient alignment and adaptively modulating their contributions enables stable and generalizable optimization on heterogeneous data.

Following this idea, a common strategy to resolve gradient conflicts is linear scalarization (LS; Fig. 1(B)), which combines losses using fixed manual weights [24]. However, LS is highly sensitive to weight selection and often requires dataset-specific re-tuning. We instead propose *Conflict-Aware Objective*

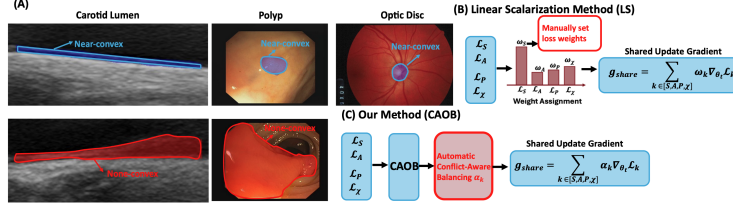


Fig. 1: (A) Examples of near-convex and non-convex anatomical structures. (B–C) Comparison between LS and CAOB in gradient weighting strategy.

Balancing (CAOB) (Fig. 1(C)). At each iteration, CAOB computes gradients of the segmentation loss and the Hadwiger prior terms, treating segmentation as the anchor objective. First, *Primary Gradient Projection (PGP)* removes only the conflicting component of each prior gradient while preserving aligned information. Then, *Adaptive Gradient Balancing (AGB)* solves normalized adaptive coefficients online to regulate the remaining prior contributions, preventing dominance from scale imbalance and ensuring stable integration of geometric priors.

Our key contributions are as follows: **1:** We propose HadBalance, a plug-and-play segmentation framework that unifies geometric shape priors and conflict-aware optimization, achieving strong generalization across diverse organs and modalities. **2:** Motivated by the prevalence of near-convex anatomy in medical data, we introduce Hadwiger-based shape priors as a unified, interpretable regularizer for near-convexity in heterogeneous medical segmentation. **3:** We propose CAOB, which resolves prior–segmentation gradient conflicts without manual weight tuning, enabling stable training on shape-heterogeneous data.

2 Methods

Preliminaries. According to Hadwiger’s characterization theorem in integral geometry, in $\mathbb{R}^2$, any valuation defined on convex bodies that is motion-invariant, additive, and continuous is a linear combination of the three intrinsic volumes [11]. In the 2D case, these correspond to the χ , P , and A , respectively. Therefore, any such valuation $\phi(\cdot)$ admits the following representation: $\phi(K) = c_0 \chi(K) + c_1 P(K) + c_2 A(K)$, where $K \subset \mathbb{R}^2$ denotes a convex body, and c_0, c_1, c_2 are constant coefficients. This motivates us to constrain these quantities as principled global shape priors, regularizing the geometric and topological properties of the prediction.

While A and P are standard geometric measures, χ is a topology-invariant scalar that captures global connectivity beyond pixel-wise accuracy [12]. Formally, for a topological space Ω , $\chi(\Omega) = \sum_{i=0}^{\infty} (-1)^i \beta_i(\Omega)$, where β_i denotes the i -th Betti number. In practice, we compute χ on discrete masks by summing local 2×2 -neighborhood contributions (enumerated by binary configurations, e.g., Gray-code indexing); see Section 2 for details [23].

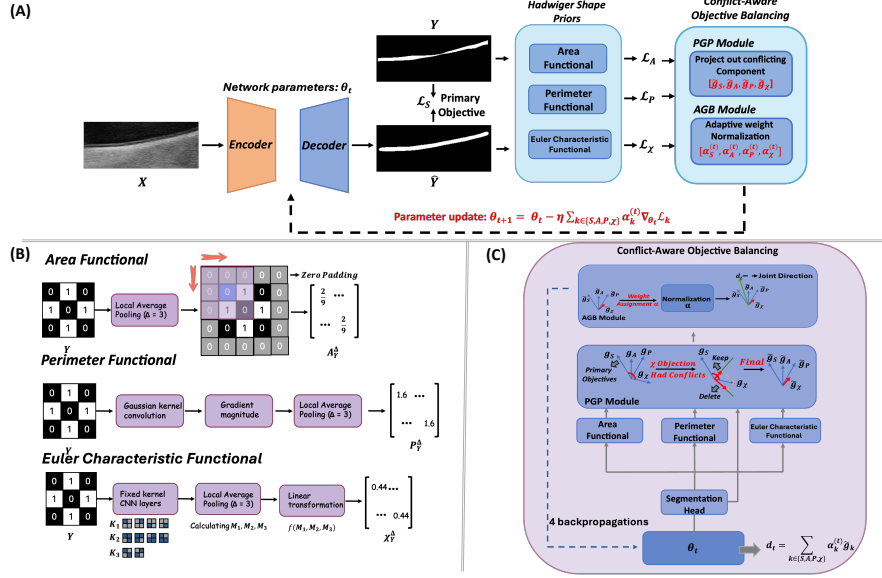


Fig. 2: (A) Overview of the framework. (B) Computation of the A , P , and χ functionals on a ground-truth example. For clarity, we illustrate local average pooling with $\Delta = 3$, while Δ is tunable in practice. (C) CAOB constructs a shared update gradient from per-loss gradients; in the figure, we illustrate the mechanism using a conflict between the χ -prior and the primary gradient.

Overall Framework. As illustrated in Fig. 2(A), given an input image X , the segmentation network f_θ produces a prediction $\hat{Y} = f_\theta(X)$. We train the network using a standard segmentation loss $\mathcal{L}_S$ augmented with Hadwiger Shape Priors regularization on A , P , and χ , implemented as MSE between the corresponding multi-scale local maps of $\hat{Y}$ and the ground truth Y . In summary, the optimization process can be formulated as

$$\theta_{t+1} = \theta_t - \eta \sum_{k \in \{S, A, P, \chi\}} \alpha_k^{(t)} \nabla_{\theta_t} \mathcal{L}_k, \quad (1)$$

where θ_t denotes the network parameters at iteration t , η is the learning rate, $\nabla_{\theta_t} \mathcal{L}_k$ denotes the gradient of $\mathcal{L}_k$ with respect to θ_t , and $\alpha_k^{(t)}$ is the weight assigned to loss term k by CAOB module.

Hadwiger Shape Priors. Incorporating Hadwiger Shape Priors requires differentiable surrogates of the discrete functionals A, P, χ on soft predictions, and it must account for their heterogeneous units and numerical ranges, which lead to mismatched gradient magnitudes during optimization. Specifically, as illustrated in Fig. 2 (B), the area functional computes multi-scale local area maps A_Y^Δ and $A_{\hat{Y}}^\Delta$ via $\Delta \times \Delta$ local average pooling (stride 1) with zero padding, and defines the

A prior loss as $\mathcal{L}_A = \sum_{\Delta \in \mathcal{D}} \text{MSE}(A_Y^\Delta, A_{\hat{Y}}^\Delta)$, where $\mathcal{D}$ denotes the set of pooling window sizes used for multi-scale computation. For the perimeter functional, we first obtain a boundary-density map by Gaussian smoothing followed by gradient magnitude (a differentiable proxy of boundary length / total variation), and then apply $\Delta \times \Delta$ average pooling (stride 1, zero padding) to obtain the local perimeter maps P_Y^Δ and $P_{\hat{Y}}^\Delta$. The P prior loss is $\mathcal{L}_P = \sum_{\Delta \in \mathcal{D}} \text{MSE}(P_Y^\Delta, P_{\hat{Y}}^\Delta)$. For the Euler characteristic, we adopt the Gray-code (bit-quad) formulation [23], which decomposes χ into a sum of local contributions over 2×2 binary neighborhoods, to compute multi-scale local χ maps for Y and $\hat{Y}$. We implement this decomposition using fixed-kernel convolutional layers that detect and count the 2×2 bit-quad configurations, and then apply $\Delta \times \Delta$ local average pooling (stride 1, zero padding) to obtain the intermediate maps $M_1^\Delta, M_2^\Delta, M_3^\Delta$. The local Euler characteristic map is finally obtained by a linear transformation: $\chi^\Delta(\cdot) = \frac{M_1^\Delta - M_2^\Delta - 2M_3^\Delta}{4}$. Accordingly, we obtain χ_Y^Δ and $\chi_{\hat{Y}}^\Delta$ for Y and $\hat{Y}$, and define the χ prior loss as $\mathcal{L}_\chi = \sum_{\Delta \in \mathcal{D}} \text{MSE}(\chi_Y^\Delta, \chi_{\hat{Y}}^\Delta)$. The Hadwiger Shape Priors regularizer is defined as

$$\mathcal{L}_H = \mathcal{L}_A + \mathcal{L}_P + \mathcal{L}_\chi. \quad (2)$$

Conflict-Aware Objective Balancing (CAOB). Unlike LS method that uses a single backward pass on a weighted-sum loss, as illustrated in Fig. 2(C), CAOB requires per-objective gradients; thus, for each mini-batch we compute $g_k = \nabla_{\theta_t} \mathcal{L}_k$ for $k \in \{S, A, P, \chi\}$ separately. PGP removes the component of g_k that opposes the segmentation gradient g_S to obtain conflict-free $\tilde{g}_k$. AGB computes $\{\alpha_k\}$ (via a lightweight non-negative least-squares fit) to balance the $\tilde{g}_k$ and form the final update direction d .

Primary Gradient Projection (PGP). Using a single shared forward pass, we obtain per-loss gradients $\{g_k\}$ via separate backpropagations. As shown in Fig. 2(C), we treat segmentation as the primary objective. For a prior-term gradient g_k , if it is negatively aligned with the segmentation gradient (i.e., $\langle g_k, g_S \rangle < 0$), it conflicts with the segmentation objective. We then remove only the component of g_k that opposes g_S , yielding a filtered gradient $\tilde{g}_k$ that satisfies $\langle \tilde{g}_k, g_S \rangle \geq 0$:

$$\tilde{g}_k = \begin{cases} g_k - \frac{\langle g_k, g_S \rangle}{\|g_S\|_2^2 + \varepsilon} g_S, & \text{if } \langle g_k, g_S \rangle < 0, \\ g_k, & \text{otherwise,} \end{cases} \quad k \in \{A, P, \chi\}, \quad (3)$$

with $\tilde{g}_S = g_S$, where ε is a constant for numerical stability. This preserves non-conflicting prior information while preventing updates that oppose segmentation.

Alpha-weight Gradient Balancing (AGB). PGP resolves directional conflicts, but heterogeneous scales can still let the remaining priors dominate. AGB computes normalized adaptive weights $\{\alpha_k\}$ to form the joint update d_t , enabling conflict-aware integration without hand-tuned loss weights. We denote the parameter increment by $\delta\theta_t$ such that $\theta_{t+1} = \theta_t + \delta\theta_t$, and apply a first-order Taylor

expansion, yielding: $\mathcal{L}_k(\theta_{t+1}) = \mathcal{L}_k(\theta_t + \delta\theta_t) \approx \mathcal{L}_k(\theta_t) + \nabla_{\theta_t} \mathcal{L}_k(\theta_t)^\top \delta\theta_t$. With gradient-based update $\theta_{t+1} = \theta_t - \eta d_t$, we have $\delta\theta_t = -\eta d_t$;

$$\mathcal{L}_k(\theta_{t+1}) - \mathcal{L}_k(\theta_t) \approx -\eta \nabla_{\theta_t} \mathcal{L}_k(\theta_t)^\top d_t = -\eta \tilde{g}_k^\top d_t. \quad (4)$$

We define $u_k^{(t)} = \tilde{g}_k^\top d_t$. If $u_k^{(t)} > 0$, $\mathcal{L}_k$ decreases, indicating a beneficial update for objective k . The joint update direction is a non-negative combination of the projected gradients, $d_t = \sum_k \alpha_k^{(t)} \tilde{g}_k$. Defining $G_t = [\tilde{g}_S, \tilde{g}_A, \tilde{g}_P, \tilde{g}_X] \in \mathbb{R}^{n \times 4}$ and $\alpha^{(t)} \in \mathbb{R}^4$ gives $d_t = G_t \alpha^{(t)}$ and $\mathbf{u}^{(t)} = G_t^\top d_t = G_t^\top G_t \alpha^{(t)}$. Based on the above, we seek a joint direction d_t that yields positive and balanced utilities across objectives. Let K denote the number of objectives (here $K = 4$). We minimize the average inverse utility $\min_{d_t} \frac{1}{K} \sum_{k=1}^K \frac{1}{u_k^{(t)}}$ which penalizes small utilities more strongly and thus encourages balanced progress. We further impose the feasibility constraints $u_k^{(t)} \geq 0$ for all k and $\|d_t\|^2 \leq \rho^2$, where ρ is a preset constant that controls the overall step scale. We formulate the above as a constrained optimization problem and introduce Lagrange multipliers $\lambda \geq 0$ and $\xi_k \geq 0$ to construct the Lagrangian:

$$\mathcal{J}(d_t, \lambda, \xi) = \frac{1}{K} \sum_{k=1}^K \frac{1}{u_k^{(t)}} + \lambda (\|d_t\|^2 - \rho^2) - \sum_{k=1}^K \xi_k u_k^{(t)}, \quad \lambda \geq 0, \xi_k \geq 0. \quad (5)$$

where λ and ξ_k are Lagrange multipliers for $\|d_t\|^2 \leq \rho^2$ and $u_k^{(t)} \geq 0$, respectively. By the Karush–Kuhn–Tucker (KKT) stationarity condition [4], the optimal direction d_t^* must satisfy $\nabla_{d_t} \mathcal{J}(d_t, \lambda, \xi) = 0$, i.e., the first-order derivative of the Lagrangian with respect to d_t vanishes. Taking the derivative w.r.t. d_t and applying complementary slackness (with $u_k^{(t)} > 0 \Rightarrow \xi_k = 0$), we obtain after simplification $\sum_{k=1}^K \frac{1}{(u_k^{(t)})^2} \tilde{g}_k = 2K\lambda d_t$. Since $2K\lambda$ is a scalar, the above equation indicates that d_t is colinear with $\sum_{k=1}^K \frac{1}{(u_k^{(t)})^2} \tilde{g}_k$. We thus write $\mathbf{u}^{(t)} = G_t^\top G_t \alpha^{(t)}$. Together with $\alpha_k^{(t)} \propto 1/(u_k^{(t)})^2$, we obtain the fixed-point equation $G_t^\top G_t \alpha^{(t)} = \sqrt{\mathbf{1}} \oslash \alpha^{(t)}$, where the square root is applied element-wise. We solve this fixed-point system for $\alpha^{(t)}$ via a constrained nonlinear least-squares solver (e.g., SciPy `least_squares` [27]), and then normalize $\alpha^{(t)}$ to remove the inherent scale ambiguity before forming the update direction $d_t = G_t \alpha^{(t)}$.

3 Experiments and Results

Dataset. We employed three representative public datasets across different organs: Carotid Ultrasound Boundary (CUBS) [18], Colonoscopy Polyps (CVC-ClinicDB) [3], and Optic Discs (DRIONS-DB) [6].

Implementation Details. In our paper, we use multi-scale pooling windows $\mathcal{D} = \{7, 15, 31\}$. For the LS baseline, all loss weights are set to unity ($w_i = 1$) for fair comparison. For CAO, we warm-start the solver at each iteration with

Table 1: Comparison with representative backbones on CUBS [18], CVC-ClinicDB [3], and DRIONS-DB [6]. The best results are highlighted in **bold**, while the best among the original backbones are denoted by double underlining.

Method	CUBS [18]			CVC-ClinicDB [3]			DRIONS-DB [6]		
	<i>DSC</i> ↑	<i>IoU</i> ↑	<i>BFScore</i> ↑	<i>DSC</i> ↑	<i>IoU</i> ↑	<i>BFScore</i> ↑	<i>DSC</i> ↑	<i>IoU</i> ↑	<i>BFScore</i> ↑
UNet [21]	85.56	75.12	70.26	92.12	86.51	68.56	93.95	88.64	61.08
nnUNet [14]	85.23	74.52	68.23	<u>93.38</u>	<u>88.55</u>	<u>73.46</u>	<u>95.79</u>	<u>92.04</u>	<u>71.60</u>
Attention-UNet [20]	85.35	74.74	70.09	92.31	86.77	69.86	94.57	89.74	66.47
UNet++ [31]	85.30	74.68	68.73	92.04	86.37	68.54	94.23	89.14	65.13
TransUNet [7]	<u>86.02</u>	<u>75.72</u>	<u>70.45</u>	91.74	86.15	66.96	94.32	89.31	66.40
SwinUNet [5]	84.96	74.16	68.39	93.54	88.21	66.20	93.83	88.43	66.97
+ NASH [19]	85.55	75.06	70.34	93.52	88.14	65.86	94.29	89.24	65.01
+ FairGrad [2]	85.38	74.92	70.45	93.49	88.06	65.05	94.23	89.14	64.53
+ LS	86.19	76.01	71.54	93.37	88.35	73.43	95.62	91.75	70.08
+ HadBalance	86.35	76.24	71.66	94.41	89.85	75.07	95.90	92.25	71.18

Table 2: Ablation study of components on CUBS [18] and CVC-ClinicDB [3].

CUBS [18]							CVC-ClinicDB [3]						
Module / Loss				Performance			Module / Loss				Performance		
L_s	Hadwiger Priors	PGP	AGB	DSC $\uparrow$	IoU $\uparrow$	BFScore $\uparrow$	L_s	Hadwiger Priors	PGP	AGB	DSC $\uparrow$	IoU $\uparrow$	BFScore $\uparrow$
✓				86.02	75.72	70.45	✓				93.38	88.55	73.46
✓	✓			86.19	76.01	71.54	✓	✓			93.37	88.35	73.43
✓	✓	✓		86.06	75.80	71.20	✓	✓	✓		93.01	88.05	73.75
✓	✓		✓	85.98	75.70	70.78	✓	✓		✓	92.25	86.72	70.73
✓	✓	✓	✓	86.35	76.23	71.66	✓	✓	✓	✓	94.41	89.85	75.07

uniform coefficients over the active losses, setting $\alpha_i^{(0)} = 1/K$. All experiments are implemented in PyTorch and run on a single NVIDIA RTX A4500 GPU.

Evaluation Metrics and Baselines. We use the Dice Similarity Coefficient (DSC) and Intersection-over-Union (IoU) to measure overlap between predictions and ground truth, and the Boundary F1-score (BFScore) to evaluate contour accuracy. We evaluate our method on multiple datasets with six representative segmentation backbones, including UNet [21], nnUNet [14], Attention-UNet [20], UNet++ [31], TransUNet [7], and SwinUNet [5]. For each dataset, we build our method on the strongest-performing backbone. To assess the proposed optimization, we further compare CAOB with Nash-MTL [19] and FairGrad [2] under identical settings using the same backbone.

Results. Table 1 summarizes a comparative evaluation of HadBalance against representative segmentation backbones and specialized gradient-optimization methods on CUBS [18], CVC-ClinicDB [3], and DRIONS-DB [6]. On CUBS [18], HadBalance improves the top-performing TransUNet baseline [7] by 0.33 in DSC, 0.52 in IoU, and 1.21 in BFScore, reaching a peak BFScore of 71.66. LS already provides a strong baseline on this near-convex dataset, achieving 86.19 DSC, 76.01 IoU, and 71.54 BFScore; HadBalance yields only marginal additional gains, likely due to limited headroom. On CVC-ClinicDB [3], which

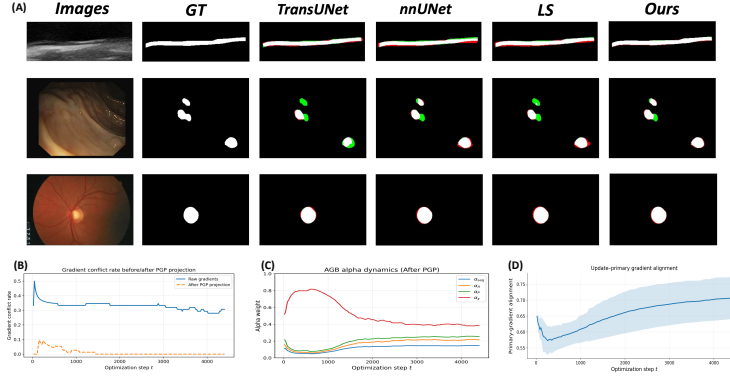


Fig. 3: (A) Qualitative results on all datasets with error overlays: white=TP, green=FN, red=FP. (B) Conflict rate between the segmentation gradient and Hadwiger gradients (A, P, χ): fraction of negative cosine similarities, before and after PGP. (C) AGB weight trajectories $\{\alpha_k\}$ after PGP. (D) Update-primary alignment (cosine similarity); shaded: rolling mean $\pm$ std.

contains more non-convex structures, HadBalance achieves clear improvements, reaching 94.41 DSC, 89.85 IoU, and 75.07 BFScore compared with 93.37 DSC, 88.35 IoU, and 73.43 BFScore from linear scalarization. Without the CAOB modules, linear weighting slightly underperforms nnUNet [14], indicating that naive loss weighting is suboptimal in non-convex settings where conflict-aware allocation becomes necessary. In addition, HadBalance outperforms NASH [19] and FairGrad [2], with the largest margin on BFScore: 75.07 compared with 65.86 for NASH [19] and 65.05 for FairGrad [2], suggesting that generic gradient balancing alone is insufficient to consistently improve boundary quality. As shown in Fig. 3(B,C), PGP markedly reduces the conflict rate and stabilizes the subsequent AGB weight dynamics; without PGP, the joint update can be dominated by geometric-prior gradients, particularly the χ term, steering optimization away from segmentation and degrading boundary quality. On DRIONS-DB [6], HadBalance achieves 95.90 DSC, 92.25 IoU, and 71.18 BFScore, improving over nnUNet by 0.11 in DSC and 0.21 in IoU. Qualitative results are shown in Fig. 3(A).

Ablation study. Table 2 reports an ablation on CUBS [18] and CVC-ClinicDB [3]. Adding Hadwiger Shape Priors improves the BCE baseline on CUBS by 0.17 DSC, 0.29 IoU, and 1.09 BFScore, but slightly reduces performance on CVC-ClinicDB by 0.01 DSC, 0.20 IoU, and 0.03 BFScore, suggesting that Hadwiger priors are sample-dependent and do not transfer without conflict-aware balancing. Using Hadwiger priors as the reference, PGP alone further reduces DSC and IoU on both datasets, indicating that removing negatively aligned prior gradients is insufficient when the remaining priors remain scale-dominant. With PGP and AGB enabled, Fig. 3(D) shows increasing alignment between the joint update direction d and the segmentation gradient g_S , and performance improves over

Hadwiger priors by 0.16 DSC, 0.22 IoU, and 0.12 BFScore on CUBS, and by 1.04 DSC, 1.50 IoU, and 1.64 BFScore on CVC-ClinicDB. Overall, CAOB combines conflict removal and adaptive balancing to enable stable use of Hadwiger priors across datasets containing both near-convex and non-convex targets.

4 Conclusion

In this paper, we are motivated by the fact that several medical segmentation targets can be approximated by globally near-convex geometry. To leverage this property, we propose **HadBalance**, which encodes near-convexity via *Hadwiger Shape Priors* (A , P , and χ) and introduces *CAOB*, a gradient-conflict-aware optimizer that adaptively balances priors against the segmentation objective on shape-heterogeneous data. Experiments on representative benchmarks show consistent improvements. Future work will reduce the additional training cost from extra backward passes.

5 Disclosure of Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Almazroa, A., Burman, R., Raahemifar, K., Lakshminarayanan, V.: Optic disc and optic cup segmentation methodologies for glaucoma image detection: a survey. *Journal of ophthalmology* **2015**(1), 180972 (2015)
2. Ban, H., Ji, K.: Fair resource allocation in multi-task learning. arXiv preprint arXiv:2402.15638 (2024)
3. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilar-íño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized medical imaging and graphics* **43**, 99–111 (2015)
4. Boyd, S., Vandenberghe, L.: *Convex optimization*. Cambridge university press (2004)
5. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: *European conference on computer vision*. pp. 205–218. Springer (2022)
6. Carmona, E.J., Rincón, M., Garcí a Feijoó, J., Martínez-de-la Casa, J.M.: Identification of the optic nerve head with genetic algorithms. *Artif. Intell. Med.* **43**(3), 243–259 (Jul 2008). <https://doi.org/10.1016/j.artmed.2008.04.005>, <http://dx.doi.org/10.1016/j.artmed.2008.04.005>
7. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
8. Fitzgerald, K., Matuszewski, B.: Fcb-swinv2 transformer for polyp segmentation. arXiv preprint arXiv:2302.01027 (2023)

9. Gao, Z., Zhou, F., Zhao, H., Chen, X., Li, X., Yu, Q., Zhao, Y., Shantsila, A., Lip, G.Y., Shantsila, E., et al.: Leveraging persistence image to enhance robustness and performance in curvilinear structure segmentation. *arXiv preprint arXiv:2601.18045* (2026)
10. Gonzalez-Jimenez, A., Lionetti, S., Gottfrois, P., Gröger, F., Navarini, A., Pouly, M.: Robust t-loss for medical image segmentation. *Medical Image Analysis* p. 103735 (2025)
11. Hadwiger, H.: *Vorlesungen über inhalt, oberfläche und isoperimetrie*, vol. 93. Springer-Verlag (2013)
12. Hatcher, A.: *Algebraic topology*. Tsinghua University Press Co., Ltd. (2005)
13. Hu, X., Li, F., Samaras, D., Chen, C.: Topology-preserving deep image segmentation. *Advances in neural information processing systems* **32** (2019)
14. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* **18**(2), 203–211 (2021)
15. Liu, J., Tai, X.C., Luo, S.: Convex shape prior for deep neural convolution network based eye fundus images segmentation. *arXiv preprint arXiv:2005.07476* (2020)
16. Liu, J., Zhang, K., Tai, X.C., Luo, S.: Convex shape prior for deep convolution neural network-based image segmentation. *Journal of Mathematical Imaging and Vision* **67**(6), 61 (2025)
17. Luo, Y., Qin, P., Chai, R., Zhai, S., Yan, J.: Boundary fusion multi-scale enhanced network for gland segmentation in colon histology images. *Biomedical Signal Processing and Control* **88**, 105566 (2024)
18. Meiburger, K.M., Zahnd, G., Faita, F., Loizou, C.P., Carvalho, C., Steinman, D.A., Gibello, L., Bruno, R.M., Marzola, F., Clarenbach, R., et al.: Carotid ultrasound boundary study (cubs): An open multicenter analysis of computerized intima-media thickness measurement systems and their clinical impact. *Ultrasound in Medicine & Biology* **47**(8), 2442–2455 (2021)
19. Navon, A., Shamsian, A., Achituve, I., Maron, H., Kawaguchi, K., Chechik, G., Fetaya, E.: Multi-task learning as a bargaining game. *arXiv preprint arXiv:2202.01017* (2022)
20. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* (2018)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
22. Royer, L.A., Richmond, D.L., Rother, C., Andres, B., Kainmueller, D.: Convexity shape constraints for image segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 402–410 (2016)
23. Savage, C.: A survey of combinatorial gray codes. *SIAM review* **39**(4), 605–629 (1997)
24. Sener, O., Koltun, V.: Multi-task learning as multi-objective optimization. *Advances in neural information processing systems* **31** (2018)
25. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. *Annual review of biomedical engineering* **19**(1), 221–248 (2017)
26. Srivastava, A., Jha, D., Chanda, S., Pal, U., Johansen, H.D., Johansen, D., Riegler, M.A., Ali, S., Halvorsen, P.: Msrf-net: A multi-scale residual fusion network for biomedical image segmentation. *IEEE Journal of Biomedical and Health Informatics* **26**(5), 2252–2263 (2021)

27. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al.: Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods* **17**(3), 261–272 (2020)
28. Wei, J., Zhao, X., Woo, J., Ouyang, J., El Fakhri, G., Chen, Q., Liu, X.: Mixture-of-shape-experts (mose): End-to-end shape dictionary framework to prompt sam for generalizable medical segmentation. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 6448–6458 (2025)
29. Yue, G., Han, W., Jiang, B., Zhou, T., Cong, R., Wang, T.: Boundary constraint network with cross layer feature integration for polyp segmentation. *IEEE Journal of Biomedical and Health Informatics* **26**(8), 4090–4099 (2022)
30. Zhou, F., Gao, Z., Zhao, H., Xie, J., Meng, Y., Zhao, Y., Lip, G.Y., Zheng, Y.: Glcp: Global-to-local connectivity preservation for tubular structure segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 237–247. Springer (2025)
31. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: *International workshop on deep learning in medical image analysis*. pp. 3–11. Springer (2018)