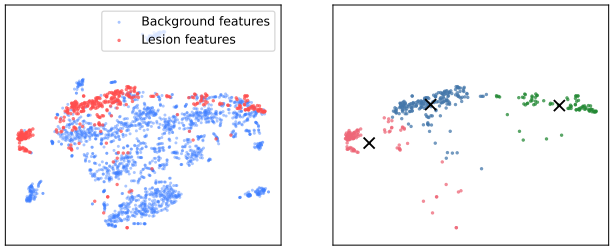
# AD-GBC: Anisotropic Granular-Ball Skip-Connection Refiner for UNet-Based Medical Image Segmentation

## Supplementary Material




(a) Background vs. lesion encoder features.

(b) Lesion-only clustering showing three distinct modes (Black "×" markers indicate the cluster centers).

Figure 7. T-SNE visualizations of encoder features extracted before AD-GBC.

## 6. Evidence of Multi-Mode Lesion Features

To support the statement made in the Introduction that *"a single semantic class (e.g., Lesion) is often not a single cluster but is multi-mode in the feature space"*, we provide additional visualizations based on encoder features extracted before the AD-GBC module.

We randomly sampled 5 images from the ISIC17 dataset along with their corresponding segmentation masks. For each image, encoder features were extracted at the pixel level and the mask was used to separate lesion and background regions. We further randomly sampled 50% of the pixels for visualization. These encoder features were flattened and projected into two dimensions using T-SNE. As shown in Fig. 7, lesion pixels do not form a single compact cluster but instead split into several distinct modes, each corresponding to different visual appearances such as dark-core areas, inflamed boundaries, or smooth lesion regions. This multi-modal behavior strongly motivates our geometric region modeling with AD-GBC.

## 7. Proof of Theorem 1

*Proof.* Let $\mathbf{C} \in \mathbb{R}^{K\times D}$ be nonzero centers and $G(\mathbf{C}) \in \mathbb{R}^{K\times K}$ their normalized Gram matrix, i.e.,

$$G(\mathbf{C})_{ij} = \frac{\mathbf{c}_i^\top \mathbf{c}_j}{\|\mathbf{c}_i\|_2 \, \|\mathbf{c}_j\|_2}, \quad i,j = 1,\dots,K.$$

By construction, $G(\mathbf{C})$ is symmetric positive semidefinite and $G(\mathbf{C})_{ii} = 1$ for all $i$, since it is the Gram matrix of unit vectors.

Assume that $\operatorname{rank}(G(\mathbf{C})) = r < K$. Because $G(\mathbf{C})$ is symmetric positive semidefinite with rank $r$, it admits a factorization

$$G(\mathbf{C}) = \mathbf{B}\mathbf{B}^\top,$$

for some $\mathbf{B} \in \mathbb{R}^{K\times r}$ with $\operatorname{rank}(\mathbf{B}) = r$ (e.g., take the top-$r$ eigenvectors of $G(\mathbf{C})$ and the square roots of the corresponding eigenvalues).

Let $\mathbf{b}_i^\top$ denote the $i$-th row of $\mathbf{B}$. Then the $(i,i)$-th entry of $G(\mathbf{C})$ satisfies

$$G(\mathbf{C})_{ii} = \left(\mathbf{B}\mathbf{B}^\top\right)_{ii} = \mathbf{b}_i^\top \mathbf{b}_i = \|\mathbf{b}_i\|_2^2.$$

Since $G(\mathbf{C})_{ii} = 1$ for all $i$, we obtain $\|\mathbf{b}_i\|_2 = 1$ for every $i = 1,\dots,K$. Hence, each row of $\mathbf{B}$ is a unit vector in $\mathbb{R}^r$.

Now define $\widetilde{\mathbf{C}} := \mathbf{B} \in \mathbb{R}^{K\times r}$ and write its $i$-th row as $\tilde{\mathbf{c}}_i^\top$. Because all rows of $\widetilde{\mathbf{C}}$ are unit-norm, its normalized Gram matrix is simply

$$G(\widetilde{\mathbf{C}})_{ij} = \tilde{\mathbf{c}}_i^\top \tilde{\mathbf{c}}_j = \mathbf{b}_i^\top \mathbf{b}_j = \left(\mathbf{B}\mathbf{B}^\top\right)_{ij} = G(\mathbf{C})_{ij},$$

for all $i, j$. Therefore,

$$G(\widetilde{\mathbf{C}}) = G(\mathbf{C}).$$

By assumption, the cosine diversity loss depends on the centers only through their normalized Gram matrix, i.e., there exists a function $\Phi : \mathbb{R}^{K\times K} \to \mathbb{R}$ such that

$$\mathcal{L}_{\text{cos-div}}(\mathbf{C}) = \Phi\big(G(\mathbf{C})\big).$$

Using $G(\widetilde{\mathbf{C}}) = G(\mathbf{C})$, we obtain

$$\mathcal{L}_{\text{cos-div}}(\widetilde{\mathbf{C}}) = \Phi\big(G(\widetilde{\mathbf{C}})\big) = \Phi\big(G(\mathbf{C})\big) = \mathcal{L}_{\text{cos-div}}(\mathbf{C}).$$

Thus, whenever $\operatorname{rank}\big(G(\mathbf{C})\big) = r < K$, there exists $\widetilde{\mathbf{C}} \in \mathbb{R}^{K\times r}$ lying in an $r$-dimensional subspace that induces exactly the same normalized Gram matrix and the same cosine diversity loss value. □

## 8. Proof of the Rank-Constrained Minimizer of the Wasserstein Uniformity Loss

In the main text, we stated that the Wasserstein uniformity loss $\mathcal{L}_{\text{W-div}}$ operates directly on the covariance matrix $\hat{\Sigma}_C$. Inspired by the formulation in Fang et al. [11], we provide here a self-contained derivation adapted to the empirical covariance setting, where $\hat{\Sigma}_C$ is constructed from a finite number of samples and is therefore rank-constrained.

**Theorem 2** (Rank-Constrained Minimizer)**.** *Let $\hat{\mu}_C \in \mathbb{R}^D$ and $\hat{\Sigma}_C \in \mathbb{R}^{D\times D}$ denote the empirical mean and covariance matrix of the centers constructed from $K$ samples. Let $\lambda_1, \ldots, \lambda_D \geq 0$ be the eigenvalues of $\hat{\Sigma}_C$, and define*

$$r := \min(D, K-1).$$

*Consider the Wasserstein uniformity loss*

$$\mathcal{L}_{\text{W-div}} = \|\hat{\mu}_C\|_2^2 + \operatorname{tr}(\hat{\Sigma}_C) - \frac{2}{\sqrt{D}} \operatorname{tr}\left(\hat{\Sigma}_C^{1/2}\right).$$

*Then the minimum of $\mathcal{L}_{\text{W-div}}$ over the feasible set is achieved when*

$$\hat{\mu}_C = \mathbf{0}, \quad \lambda_1 = \cdots = \lambda_r = \frac{1}{D}, \quad \lambda_{r+1} = \cdots = \lambda_D = 0.$$

*Proof.* Since $\hat{\Sigma}_C$ is symmetric positive semidefinite, all its eigenvalues satisfy $\lambda_k \geq 0$ for $k = 1, \ldots, D$. Moreover, because $\hat{\Sigma}_C$ is constructed from $K$ samples, its rank is constrained by

$$m := \operatorname{rank}(\hat{\Sigma}_C) \leq r := \min(D, K-1).$$

We can assume without loss of generality that the first $m$ eigenvalues are strictly positive, while the remaining $D-m$ eigenvalues are zero.

By spectral calculus, the trace operations only depend on the $m$ nonzero eigenvalues:

$$\operatorname{tr}(\hat{\Sigma}_C) = \sum_{k=1}^{m} \lambda_k, \qquad \operatorname{tr}\left(\hat{\Sigma}_C^{1/2}\right) = \sum_{k=1}^{m} \sqrt{\lambda_k}.$$

Therefore, the loss can be rewritten as a sum over these $m$ eigenvalues:

$$\mathcal{L}_{\text{W-div}} = \|\hat{\mu}_C\|_2^2 + \sum_{k=1}^{m} \left(\lambda_k - \frac{2}{\sqrt{D}}\sqrt{\lambda_k}\right).$$

For each $k \in \{1, \ldots, m\}$, define $x_k := \sqrt{\lambda_k} > 0$. Completing the square yields

$$\lambda_k - \frac{2}{\sqrt{D}}\sqrt{\lambda_k} = x_k^2 - \frac{2}{\sqrt{D}}x_k = \left(x_k - \frac{1}{\sqrt{D}}\right)^2 - \frac{1}{D}.$$

Substituting this identity into the loss gives

$$\mathcal{L}_{\text{W-div}} = \|\hat{\mu}_C\|_2^2 + \sum_{k=1}^{m} \left(x_k - \frac{1}{\sqrt{D}}\right)^2 - \frac{m}{D}.$$

For any feasible $\hat{\Sigma}_C$, the first two terms are sums of nonnegative values. To minimize the overall loss, we must make the subtracted term $\frac{m}{D}$ as large as possible, which requires maximizing the rank $m$ up to its upper bound $r$. Thus, the global minimum over the feasible set is achieved when

$$\hat{\mu}_C = \mathbf{0}, \qquad m = r, \qquad x_k = \frac{1}{\sqrt{D}} \text{ for all } k = 1, \ldots, r.$$

Equivalently,

$$\lambda_1 = \cdots = \lambda_r = \frac{1}{D}, \qquad \lambda_{r+1} = \cdots = \lambda_D = 0.$$

Therefore, the minimum is achieved when the feasible rank reaches its maximum ($r$), and all nonzero eigenvalues are equal to ($1/D$). □

This corresponds to a spectrally uniform distribution within the feasible rank-r subspace, which promotes a well-spread set of centers without contradicting the anisotropic modeling of individual regions.

## 9. Detailed Complexity and Efficiency Comparison of the AD-GBC Module

We provide a detailed breakdown of the complexity analysis presented in Section 3.5. of the main paper. Let the input feature map have shape $B \times C_{in} \times H \times W$, the number of pixels $N = H \times W$, the number of granular balls be $K$, and the projection dimension be $D$. We assume $B = 1$ for this analysis, as FLOPs scale linearly with batch size.

**Computational Complexity (FLOPs)** We sum the FLOPs for each step of the forward pass, as shown in Fig. 2 in the main paper:

- *Projection Layers ($f_{proj}$, $\tilde{f}_{proj}$):* The initial $1 \times 1$ Conv ($C_{in} \to D$) requires $\mathcal{O}(N \cdot C_{in} \cdot D)$ FLOPs. The final $1 \times 1$ Conv ($D \to C_{in}$) also requires $\mathcal{O}(N \cdot D \cdot C_{in})$ FLOPs. The total cost for projections is $\mathcal{O}(N \cdot C_{in} \cdot D)$.
- *D-GBC Core Interaction:* This is the main component.
  1) *Scaled Distance ($d_{i,k}$):* Calculating the squared scaled Euclidean distance $\|(\mathbf{z}_i - \mathbf{c}_k) \oslash \boldsymbol{\sigma}_k\|_2^2$. This operation is performed for all $N$ pixels against all $K$ balls, each in $D$ dimensions. This involves subtractions, divisions, squaring, and summing, resulting in $\mathcal{O}(N \cdot K \cdot D)$ FLOPs.
  2) *Softmax ($\alpha_{i,k}$):* Normalizing the $N \times K$ distance matrix requires $\mathcal{O}(N \cdot K)$ operations, which is negligible compared to the $\mathcal{O}(N \cdot K \cdot D)$ steps.
  3) *Aggregation (Set→Ball, $\mathbf{c}'_k$):* This is a weighted sum $\sum_i \alpha_{i,k}\mathbf{z}_i$, equivalent to a matrix multiplication $\alpha^T\mathbf{Z}$ ($\mathbb{R}^{K\times N} \times \mathbb{R}^{N\times D}$). This requires $\mathcal{O}(N \cdot K \cdot D)$ FLOPs.
  4) *Broadcast (Ball→Set, $\hat{\mathbf{z}}_i$):* This is a weighted sum $\sum_k \alpha_{i,k}\mathbf{c}'_k$, equivalent to $\alpha\mathbf{C}'$ ($\mathbb{R}^{N\times K} \times \mathbb{R}^{K\times D}$). This also requires $\mathcal{O}(N \cdot K \cdot D)$ FLOPs.

  Therefore, the total complexity of the core interaction is $\mathcal{O}(N \cdot K \cdot D)$.

- *Refinement Block ($f_{refine}$):* The $3 \times 3$ depthwise convolution ($C_{in} \rightarrow C_{in}$) requires FLOPs proportional to the output size ($N \cdot C_{in}$) and the kernel operations ($C_{in} \cdot 3 \cdot 3$). This results in $\mathcal{O}(N \cdot C_{in}^2)$ FLOPs.

Summing these components, the total computational complexity is:

$$\mathcal{O}(N \cdot C_{in} \cdot D + N \cdot K \cdot D + N \cdot C_{in}^2) = \mathcal{O}(N \cdot (C_{in} D + K D + C_{in}^2)).$$

Since $C_{in}$, $K$, and $D$ are fixed hyperparameters, the module's computational cost scales linearly with the number of pixels $N$.

**Parameter Complexity**

- *GB Anchors:* The $K$ centers ($\mathbf{c}_k \in \mathbb{R}^D$) and $K$ anisotropic scales ($\boldsymbol{\sigma}_k \in \mathbb{R}^D$) require $K \cdot D + K \cdot D = 2KD$ parameters.
- *Optional Projection Layers:* The forward projection $f_{\text{proj}}$ ($1\times1$ Conv ($C_{in} \rightarrow D$) + BN) uses $C_{in}D$ convolution weights and $2D$ BN parameters. The backward projection uses $DC_{in}$ convolution weights and $2C_{in}$ BN parameters.
- *Refinement Block:* The $f_{\text{refine}}$ block ($3 \times 3$ Conv ($C_{in} \rightarrow C_{in}$) + BN) requires $(C_{in} \cdot C_{in} \cdot 9)$ weights and $2C_{in}$ (scale/shift) parameters from BN.

The total number of parameters is $\mathcal{O}(KD + C_{in}D + C_{in}^2)$. This count is fixed once the hyperparameters are set and does not grow with $N$.

# 10. Additional Visualizations and Diagnostic Analyses

## 10.1. Extended Qualitative Comparisons

This section provides extended qualitative comparisons to complement the main paper. Fig. 8 shows a larger set of visual examples across four datasets in a row-wise manner: BUSI (breast ultrasound), GlaS (colon histology), CVC-ClinicDB (colonoscopy), and ISIC17 (dermoscopy). BUSI is also visualized in the main paper (Fig. 5); here we include additional three datasets to more comprehensively assess the generality of our method. For each dataset, we show the input image with its ground-truth (GT) contour, followed by predictions from seven competing methods. These extended examples further validate the findings reported in the main paper: the AD-GBC module provides more stable, geometry-aware region modeling, leading to visibly more accurate and consistent segmentation boundaries across diverse medical imaging modalities.

Across datasets, models equipped with the proposed AD-GBC module consistently show reduced purple and yellow bands along ambiguous or irregular boundaries, indicating fewer over- and under-segmentation artifacts. In contrast, baseline models without AD-GBC frequently exhibit jagged, incomplete, or overly expanded contours, particularly in cases with low contrast (BUSI), tightly packed glands (GlaS), elongated polyp structures (CVC-ClinicDB), or fuzzy lesion borders (ISIC17).

## 10.2. Diagnostic Visualization of AD-GBC Center Activations

Fig. 9 presents a diagnostic visualization of all learned AD-GBC centers (K=32) for one representative ISIC17 image. These maps are not used for qualitative comparison, but serve as an internal inspection of the activation patterns learned by the AD-GBC module.

It is important to note that the centers are not designed to form interpretable "parts" individually. Because the AD-GBC module uses soft-assignment weighting and multi-region aggregation, most centers specialize in sub-regions or intermediate feature patterns, rather than forming complete semantic shapes. Therefore, only a subset of centers exhibit clear lesion-aligned responses, while others highlight texture, chromatic variation, or transitional boundary structures. This behavior is expected and consistent with our multi-center design.

These visualizations are provided solely to illustrate the diversity and specialization of center responses. The final segmentation quality arises from the joint aggregation of all centers, not from the interpretability of any single map. Thus, while individual centers may appear noisy or partial, their combined effect yields the stable and geometry-aware boundaries demonstrated in the main paper.

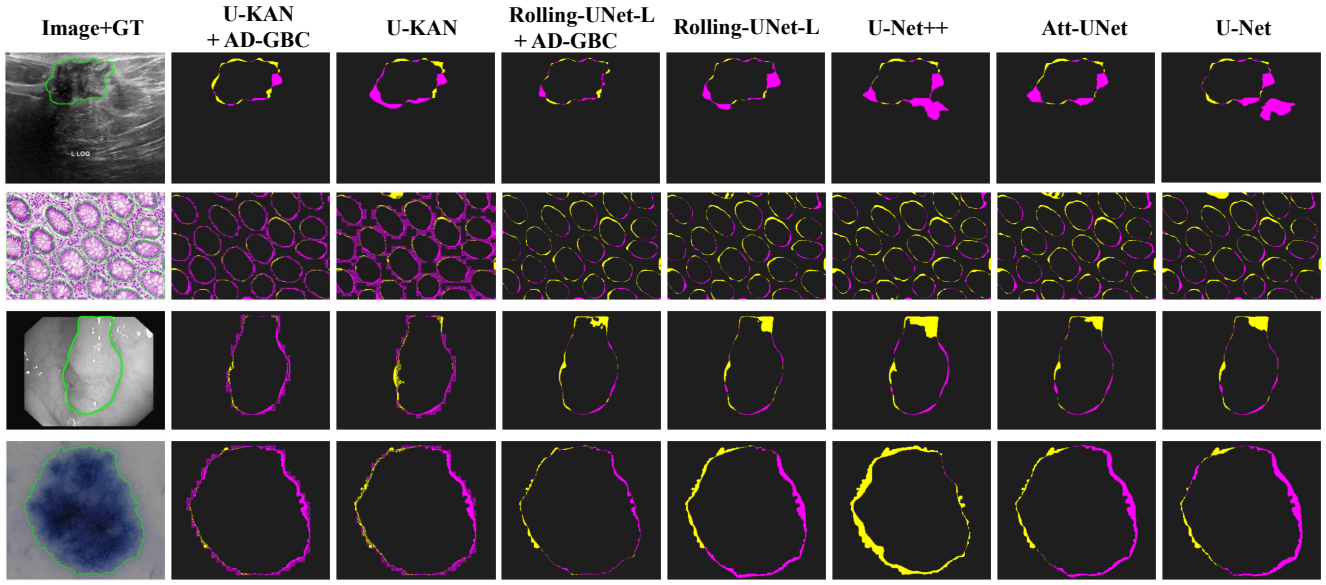



Figure 8. Additional qualitative comparisons across datasets. From top to bottom are the BUSI, GlaS, CVC-ClinicDB and ISIC17 datasets. The first column is the original image, with the green contour indicating the Ground Truth. In the visualized segmentation results, purple indicates over-segmentation, and yellow indicates under-segmentation.

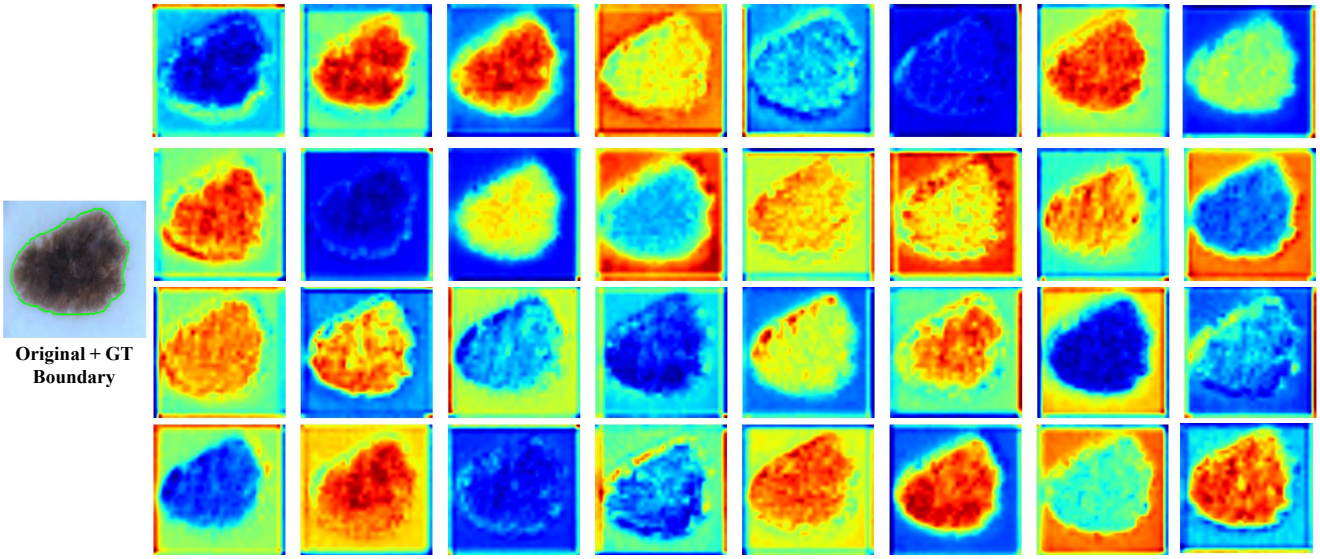



Figure 9. Diagnostic visualization of all 32 AD-GBC centers for a representative ISIC17 image. The first column shows the original image with its ground-truth boundary, followed by the activation maps of all AD-GBC centers. These visualizations are provided for diagnostic analysis and completeness.