

# RobustMedSAM: Degradation-Resilient Medical Image Segmentation via Robust Foundation Model Adaptation

## Supplementary Material

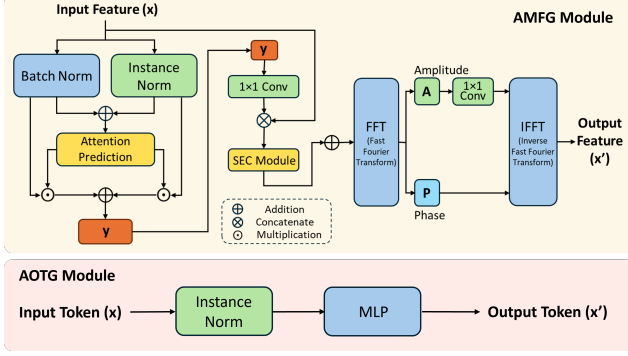


Figure S1. Detailed architecture of Anti-degradation Mask Feature Generation and Anti-degradation Output Token Generation modules.

### S1. Evaluation Metrics

Let  $P$  denote the predicted mask and  $G$  the ground-truth mask.

#### Dice Coefficient

$$\text{Dice}(P, G) = \frac{2|P \cap G|}{|P| + |G|}. \quad (\text{S1})$$

#### Intersection over Union (IoU)

$$\text{IoU}(P, G) = \frac{|P \cap G|}{|P \cup G|}. \quad (\text{S2})$$

**Normalized Surface Distance (NSD)** Let  $\partial P$  and  $\partial G$  denote the surfaces of  $P$  and  $G$ . Given tolerance  $\tau$ , NSD is defined as

$$\text{NSD}(P, G) = \frac{|\{x \in \partial P : d(x, \partial G) \leq \tau\}| + |\{y \in \partial G : d(y, \partial P) \leq \tau\}|}{|\partial P| + |\partial G|} \quad (\text{S3})$$

where  $d(\cdot, \partial G)$  denotes the shortest Euclidean distance to surface  $\partial G$ .

### S2. Loss Weight Selection

We set loss weights to balance term magnitude and functional role:  $\alpha=20$  compensates for severe foreground-background imbalance and emphasizes hard pixels, while  $\beta=1$  keeps overlap optimization on a stable scale.

Since decoder-side MSE regularizers are typically low-magnitude, we set  $\lambda_1=100$  as the primary constraint for degradation-invariant decoding and  $\lambda_2=2$  as a token-level stabilizer to avoid suppressing  $\mathcal{L}_{\text{seg}}$ .

### S3. Architecture Details

#### S3.1. Anti-degradation Mask Feature Generation

We adopt the Anti-degradation Mask Feature Generation module proposed in [9] to enhance feature robustness under image degradation. Given an input feature map  $F$ , the module first processes features using two parallel normalization branches. Instance Normalization (IN) is used to reduce degradation-related style variations, while Batch Normalization (BN) preserves structural details that may be weakened by IN. The outputs of the two branches are fused and refined using an attention mechanism to adaptively balance their contributions. The fused features are concatenated with the original input along the channel dimension and further refined by a channel attention module. To suppress degradation patterns, the features are transformed into the frequency domain using the Fourier transform. The amplitude component captures degradation-related style information and is filtered using a  $1 \times 1$  convolution, while the phase component preserves structural information. The refined features are then transformed back to the spatial domain via the inverse Fourier transform.

#### S3.2. Anti-degradation Output Token Generation

To enhance robustness against image degradation, we adopt an Anti-Degradation Output Token Generation module proposed in [9] to refine the Robust Output Token per mask ( $T_{RO}$ ). Compared with mask features,  $T_{RO}$  mainly encodes classification boundary information and contains limited texture details. Therefore, a lightweight refinement module is sufficient to suppress degradation-sensitive information. Specifically, the token is processed through several layers of Instance Normalization followed by a single MLP layer. Instance Normalization reduces degradation-related style variations, while the MLP further refines the token representation with minimal computational overhead.

The refined token is finally combined with the Robust Mask Feature to produce the final segmentation mask.

Table S1. Clean vs. degraded with **point prompts** (Dice  $\uparrow$ ).  $\Delta$  = degraded – clean; positive values indicate the model benefits from the clear-degraded pair protocol.

Dataset	SAM			MedSAM			RobustMedSAM			+SVD		
	Clean	Deg.	$\Delta$	Clean	Deg.	$\Delta$	Clean	Deg.	$\Delta$	Clean	Deg.	$\Delta$
ISIC 2016	0.734	0.752	+0.018	0.203	0.202	-0.001	0.809	0.911	+0.102	0.699	0.910	+0.210
BUSI	0.717	0.718	+0.001	0.252	0.249	-0.002	0.610	0.760	+0.151	0.434	0.678	+0.243
PROMISE12	0.639	0.593	-0.046	0.208	0.209	+0.001	0.714	0.805	+0.091	0.544	0.751	+0.207
UWaterloo	0.840	0.841	+0.001	0.201	0.202	+0.002	0.736	0.869	+0.132	0.759	0.872	+0.113
BriFiSeg	0.321	0.311	-0.010	0.112	0.114	+0.002	0.289	0.305	+0.016	0.125	0.436	+0.311
AbdomenUS	0.477	0.463	-0.013	0.227	0.227	+0.001	0.563	0.666	+0.103	0.261	0.612	+0.351
<b>Overall</b>	0.621	0.613	<b>-0.008</b>	0.200	0.201	<b>+0.000</b>	0.620	0.719	<b>+0.099</b>	0.470	0.710	<b>+0.239</b>

Table S2. Per-degradation with **point prompts** — Dice  $\uparrow$ . Values are mean  $\pm$  std across datasets. (Single-dataset degradations show mean only.)

Degradation	SAM	MedSAM	RobustSAM	Ours - Robust MedSAM	Ours - Robust MedSAM+SVD
Brightness	0.619 $\pm 0.177$	0.206 $\pm 0.040$	0.633 $\pm 0.196$	0.729 $\pm 0.195$	0.710 $\pm 0.166$
Color Jitter	0.801 $\pm 0.059$	0.201 $\pm 0.017$	0.870 $\pm 0.043$	0.897 $\pm 0.012$	0.880 $\pm 0.024$
Compression	0.813 $\pm 0.042$	0.198 $\pm 0.002$	0.879 $\pm 0.042$	0.887 $\pm 0.024$	0.905 $\pm 0.004$
Contrast	0.605 $\pm 0.176$	0.189 $\pm 0.054$	0.641 $\pm 0.200$	0.717 $\pm 0.182$	0.696 $\pm 0.167$
Gaussian Blur	0.609 $\pm 0.163$	0.203 $\pm 0.050$	0.632 $\pm 0.206$	0.711 $\pm 0.213$	0.705 $\pm 0.169$
Gaussian Noise	0.625 $\pm 0.213$	0.206 $\pm 0.041$	0.639 $\pm 0.216$	0.717 $\pm 0.230$	0.703 $\pm 0.182$
Poisson	0.321	0.123	0.334	0.345	0.472
Rayleigh	0.665	0.214	0.563	0.816	0.772
Rician	0.591	0.203	0.593	0.810	0.773
Salt & Pepper	0.582 $\pm 0.095$	0.257 $\pm 0.004$	0.525 $\pm 0.042$	0.701 $\pm 0.056$	0.656 $\pm 0.038$
Speckle	0.594 $\pm 0.128$	0.217 $\pm 0.005$	0.602 $\pm 0.128$	0.717 $\pm 0.041$	0.652 $\pm 0.049$
Step Motion MRI	0.453	0.205	0.464	0.779	0.712

## S4. Additional Experimental Results

## S4.2. Detailed Per-Degradation Analysis

### S4.1. Performance on Clear vs Degraded Images

Table S1 reports segmentation performance under clean and degraded imaging conditions. The vanilla models (SAM and MedSAM) show negligible or negative changes overall, indicating limited robustness to degraded inputs. In contrast, the robust variants, particularly the SVD-based method, consistently benefit from the clear-degraded pair protocol and achieve substantial improvements across datasets, with the largest overall gain of +0.240.

Tables S2 S3 S4 presents the Dice, IoU and NSD of all models under different degradation types. Overall, the robust variants consistently outperform the vanilla SAM and MedSAM across most degradations, indicating improved robustness to diverse image corruptions. In particular, the SVD-enhanced variant shows notable advantages for certain noise-related degradations (e.g., compression and Poisson noise), while RobustMedSAM achieves strong and stable performance across most other degradation types.

Table S3. Per-degradation analysis with **point prompts** — IoU  $\uparrow$ . Values are mean  $\pm$  std across datasets. (Single-dataset degradations show mean only.)

Degradation	SAM	MedSAM	RobustSAM	Ours - Robust MedSAM	Ours - Robust MedSAM+SVD
Brightness	0.506 $\pm 0.190$	0.123 $\pm 0.027$	0.523 $\pm 0.219$	0.621 $\pm 0.212$	0.595 $\pm 0.192$
Color Jitter	0.704 $\pm 0.079$	0.120 $\pm 0.013$	0.787 $\pm 0.058$	0.820 $\pm 0.020$	0.799 $\pm 0.035$
Compression	0.709 $\pm 0.053$	0.118 $\pm 0.003$	0.802 $\pm 0.057$	0.809 $\pm 0.033$	0.832 $\pm 0.010$
Contrast	0.494 $\pm 0.187$	0.113 $\pm 0.036$	0.532 $\pm 0.224$	0.607 $\pm 0.199$	0.582 $\pm 0.193$
Gaussian Blur	0.494 $\pm 0.175$	0.120 $\pm 0.033$	0.522 $\pm 0.227$	0.605 $\pm 0.221$	0.591 $\pm 0.193$
Gaussian Noise	0.519 $\pm 0.226$	0.123 $\pm 0.027$	0.532 $\pm 0.235$	0.615 $\pm 0.232$	0.589 $\pm 0.201$
Poisson	0.195	0.068	0.205	0.224	0.324
Rayleigh	0.552	0.127	0.441	0.703	0.647
Rician	0.480	0.120	0.460	0.693	0.645
Salt & Pepper	0.458 $\pm 0.109$	0.160 $\pm 0.006$	0.401 $\pm 0.055$	0.571 $\pm 0.069$	0.526 $\pm 0.049$
Speckle	0.473 $\pm 0.140$	0.131 $\pm 0.004$	0.479 $\pm 0.140$	0.584 $\pm 0.054$	0.521 $\pm 0.059$
Step Motion MRI	0.341	0.121	0.338	0.652	0.577

Table S4. Per-degradation analysis with **point prompts** — NSD  $\uparrow$  ( $\tau=2$ px). Values are mean  $\pm$  std across datasets. (Single-dataset degradations show mean only.)

Degradation	SAM	MedSAM	RobustSAM	Ours - Robust MedSAM	Ours - Robust MedSAM+SVD
Brightness	0.191 $\pm 0.088$	0.036 $\pm 0.011$	0.203 $\pm 0.112$	0.199 $\pm 0.066$	0.182 $\pm 0.053$
Color Jitter	0.212 $\pm 0.138$	0.038 $\pm 0.021$	0.274 $\pm 0.149$	0.262 $\pm 0.028$	0.245 $\pm 0.022$
Compression	0.228 $\pm 0.107$	0.030 $\pm 0.012$	0.296 $\pm 0.133$	0.275 $\pm 0.023$	0.252 $\pm 0.019$
Contrast	0.184 $\pm 0.092$	0.036 $\pm 0.011$	0.209 $\pm 0.127$	0.205 $\pm 0.069$	0.195 $\pm 0.062$
Gaussian Blur	0.178 $\pm 0.080$	0.037 $\pm 0.010$	0.195 $\pm 0.105$	0.202 $\pm 0.066$	0.194 $\pm 0.060$
Gaussian Noise	0.190 $\pm 0.116$	0.038 $\pm 0.012$	0.207 $\pm 0.141$	0.199 $\pm 0.069$	0.193 $\pm 0.058$
Poisson	0.108	0.051	0.104	0.143	0.193
Rayleigh	0.192	0.029	0.132	0.218	0.188
Rician	0.170	0.028	0.132	0.206	0.189
Salt & Pepper	0.153 $\pm 0.041$	0.041 $\pm 0.003$	0.165 $\pm 0.002$	0.170 $\pm 0.061$	0.157 $\pm 0.058$
Speckle	0.176 $\pm 0.053$	0.041 $\pm 0.002$	0.215 $\pm 0.022$	0.165 $\pm 0.061$	0.154 $\pm 0.066$
Step Motion MRI	0.091	0.027	0.078	0.164	0.155

### S4.3. Segmentation Performance Under Box Prompt

Table S5 presents segmentation performance under bounding box prompts across multiple image degradation condi-

tions. Results are reported as the mean  $\pm$  standard deviation over different degradation types for each dataset. We compare SAM, MedSAM, and Robust MedSAM using Dice and IoU metrics on six medical imaging datasets. MedSAM

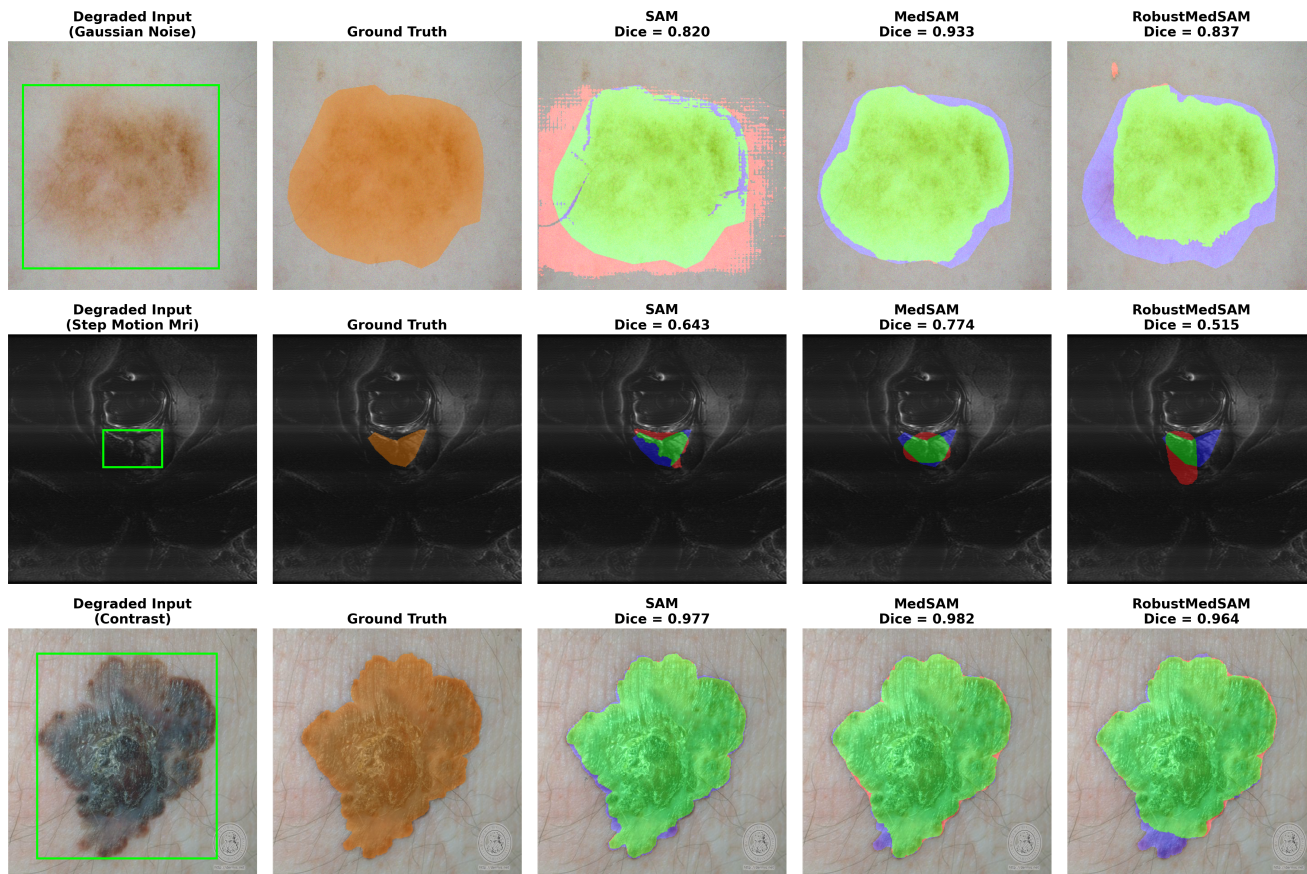


Figure S2. **Bounding Box:** Qualitative segmentation results under degradation. Each panel shows (left to right): degraded input with "bounding box prompts, ground truth, SAM, MedSAM, RobustMedSAM predictions. Green = true positive, red = false positive, blue = false negative.

consistently outperforms SAM, indicating the effectiveness of medical-domain adaptation. Robust MedSAM further improves robustness under degraded conditions, achieving the best overall Dice score and demonstrating improved stability across challenging datasets.

Figure S2 shows qualitative segmentation results under different image degradations using bounding box prompts. Each row corresponds to a representative degradation type, including Gaussian noise, motion artifacts, and contrast variation. From left to right: degraded input with bounding box prompt, ground truth, and predictions from SAM, MedSAM, and Robust MedSAM. Green indicates true positives, red indicates false positives, and blue indicates false negatives. Overall, MedSAM and Robust MedSAM produce more accurate and stable segmentations than SAM under degraded conditions.

Table S5. **Bounding box prompt** segmentation performance under various image degradations. Values are reported as mean  $\pm$  std across degradation types per dataset.

Dataset	Dice $\uparrow$			IoU $\uparrow$		
	SAM	MedSAM	Robust MedSAM	SAM	MedSAM	Robust MedSAM
ISIC 2016	0.851 ( $\pm 0.012$ )	0.929 ( $\pm 0.003$ )	0.910 ( $\pm 0.006$ )	0.747 ( $\pm 0.018$ )	0.869 ( $\pm 0.004$ )	0.844 ( $\pm 0.009$ )
BUSI	0.778 ( $\pm 0.019$ )	0.911 ( $\pm 0.013$ )	0.807 ( $\pm 0.012$ )	0.683 ( $\pm 0.020$ )	0.840 ( $\pm 0.020$ )	0.701 ( $\pm 0.012$ )
PROMISE12	0.876 ( $\pm 0.038$ )	0.886 ( $\pm 0.018$ )	0.888 ( $\pm 0.009$ )	0.793 ( $\pm 0.048$ )	0.803 ( $\pm 0.028$ )	0.804 ( $\pm 0.012$ )
UWaterloo Skin Cancer	0.881 ( $\pm 0.011$ )	0.956 ( $\pm 0.007$ )	0.869 ( $\pm 0.017$ )	0.797 ( $\pm 0.017$ )	0.917 ( $\pm 0.012$ )	0.791 ( $\pm 0.019$ )
BriFiSeg	0.250 ( $\pm 0.021$ )	0.244 ( $\pm 0.007$ )	0.505 ( $\pm 0.035$ )	0.146 ( $\pm 0.013$ )	0.142 ( $\pm 0.005$ )	0.348 ( $\pm 0.031$ )
AbdomenUS	0.405 ( $\pm 0.113$ )	0.774 ( $\pm 0.008$ )	0.782 ( $\pm 0.006$ )	0.307 ( $\pm 0.096$ )	0.654 ( $\pm 0.010$ )	0.673 ( $\pm 0.006$ )
<b>Overall</b>	0.691 ( $\pm 0.251$ )	0.801 ( $\pm 0.234$ )	0.804 ( $\pm 0.131$ )	0.597 ( $\pm 0.255$ )	0.723 ( $\pm 0.251$ )	0.706 ( $\pm 0.159$ )

Table S6. Per-degradation analysis (**bounding box prompt**). Values are mean  $\pm$  std across datasets.

Degradation	Dice $\uparrow$			IoU $\uparrow$		
	SAM	MedSAM	Robust MedSAM	SAM	MedSAM	Robust MedSAM
Brightness	0.671 $\pm 0.275$	0.788 $\pm 0.277$	0.790 $\pm 0.151$	0.575 $\pm 0.281$	0.713 $\pm 0.297$	0.689 $\pm 0.183$
Color Jitter	0.862 $\pm 0.016$	0.944 $\pm 0.020$	0.868 $\pm 0.046$	0.766 $\pm 0.029$	0.895 $\pm 0.034$	0.795 $\pm 0.052$
Compression	0.885 $\pm 0.013$	0.943 $\pm 0.022$	0.898 $\pm 0.026$	0.799 $\pm 0.023$	0.894 $\pm 0.039$	0.829 $\pm 0.035$
Contrast	0.654 $\pm 0.271$	0.786 $\pm 0.276$	0.798 $\pm 0.138$	0.557 $\pm 0.275$	0.710 $\pm 0.296$	0.696 $\pm 0.171$
Gaussian Blur	0.647 $\pm 0.292$	0.785 $\pm 0.266$	0.800 $\pm 0.150$	0.552 $\pm 0.290$	0.705 $\pm 0.284$	0.701 $\pm 0.183$
Gaussian Noise	0.702 $\pm 0.267$	0.782 $\pm 0.270$	0.785 $\pm 0.172$	0.607 $\pm 0.276$	0.702 $\pm 0.288$	0.686 $\pm 0.199$
Poisson	0.233 $\pm 0.101$	0.240 $\pm 0.092$	0.535 $\pm 0.123$	0.136 $\pm 0.064$	0.139 $\pm 0.059$	0.375 $\pm 0.113$
Rayleigh	0.915 $\pm 0.042$	0.894 $\pm 0.077$	0.899 $\pm 0.061$	0.846 $\pm 0.068$	0.816 $\pm 0.115$	0.822 $\pm 0.084$
Rician	0.909 $\pm 0.046$	0.882 $\pm 0.076$	0.894 $\pm 0.063$	0.836 $\pm 0.074$	0.796 $\pm 0.114$	0.813 $\pm 0.084$
Salt & Pepper	0.668 $\pm 0.195$	0.842 $\pm 0.078$	0.796 $\pm 0.019$	0.567 $\pm 0.204$	0.743 $\pm 0.104$	0.689 $\pm 0.018$
Speckle	0.557 $\pm 0.329$	0.836 $\pm 0.085$	0.787 $\pm 0.013$	0.465 $\pm 0.321$	0.737 $\pm 0.113$	0.678 $\pm 0.011$
Step Motion MRI	0.836 $\pm 0.117$	0.849 $\pm 0.089$	0.888 $\pm 0.067$	0.733 $\pm 0.154$	0.748 $\pm 0.125$	0.803 $\pm 0.085$