SAMWave: Adapting Segment Anything Model to Difficult Tasks

INDRAPRASTHA INSTITUTE of INFORMATION TECHNOLOGY DELHI

Saurabh Yadav, Avi Gupta*, Koteswar Rao Jerripothula*

{saurabhy, avig}@iiitd.ac.in, kotesrj@iitk.ac.in

BMVC 2025

MOTIVATION

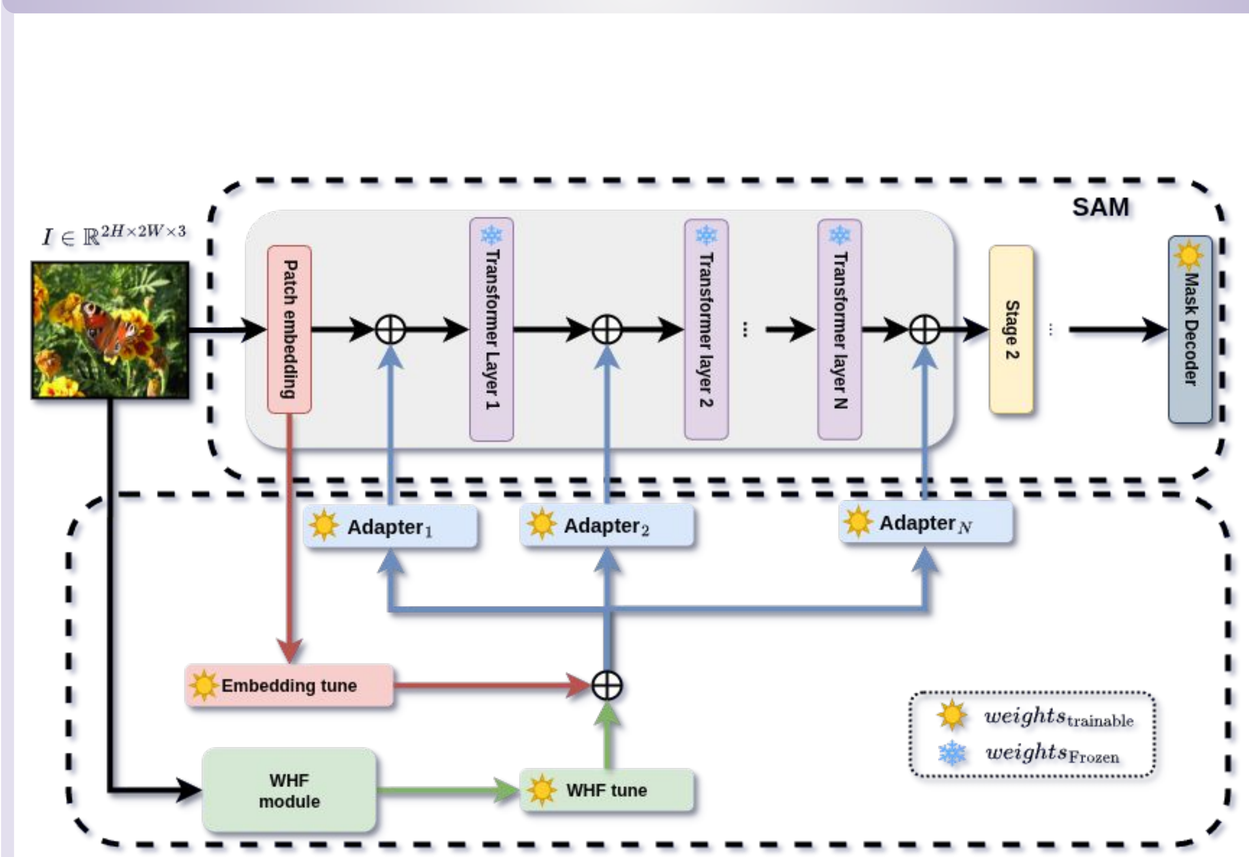
- For difficult images such as camouflaged images, polyp images etc, SAM is unable to perform optimally due to difficult boundaries.
- Previous fine-tuning methods use high-frequency features for training on such images.
- However, Analysis reveals that these approaches offer limited benefits due to constraints in their feature extraction techniques.

High Frequency features extracted from other methods give similar results when low-frequency features are provided.

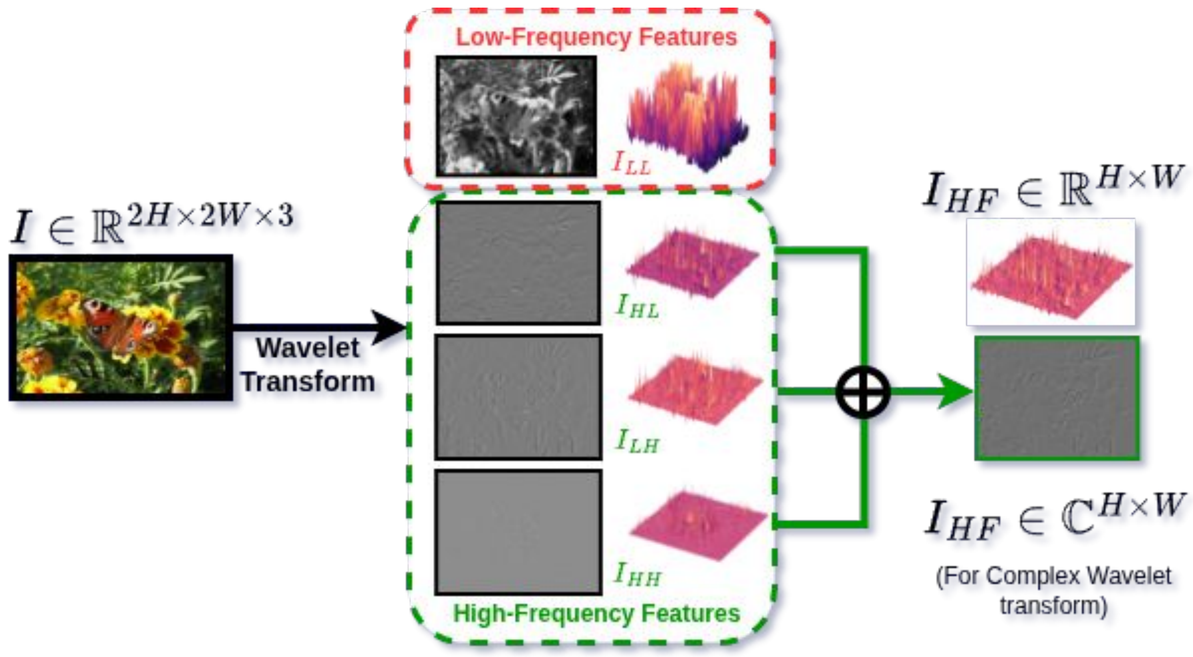
	CHAMELEON				CAMO				COD10K			
	$\mathcal{S}_m\uparrow$	$\mathcal{E}_\phi\uparrow$	$\mathcal{F}_\beta^w\uparrow$	$\mathcal{M}\downarrow$	$\mathcal{S}_m\uparrow$	$\mathcal{E}_\phi\uparrow$	$\mathcal{F}_\beta^w\uparrow$	$\mathcal{M}\downarrow$	$\mathcal{S}_m\uparrow$	$\mathcal{E}_\phi\uparrow$	$\mathcal{F}_\beta^w\uparrow$	$\mathcal{M}\downarrow$
HFreq	**0.896**	**0.919**	0.824	**0.033**	**0.847**	**0.873**	**0.765**	**0.070**	**0.883**	**0.918**	0.801	**0.025**
LFreq	0.888	0.903	**0.830**	0.034	0.824	0.843	**0.765**	0.078	0.872	0.901	**0.803**	0.027
	+ 0.008	+ 0.016	- 0.006	+ 0.001	+ 0.023	+ 0.030	0.000	+ 0.008	+ 0.009	+ 0.017	- 0.002	+ 0.002

To overcome this, we propose an efficient approach that utilizes the wavelet transform to extract richer high-frequency features from input data.

SAMWave



Wavelet-based Adapters



- Using wavelet transform provides multi-fold advantage:
  - Both frequency and spatial position of frequencies can be extracted.
  - High-frequency features are image-dependent and are easily separable from low-frequency features.
  - Near shift-invariant features can be obtained using Complex wavelet transform.

$\mathbb{R}$-Adapters

$$Emb_{real} = f_{real}(I_{HF})$$

Real-valued

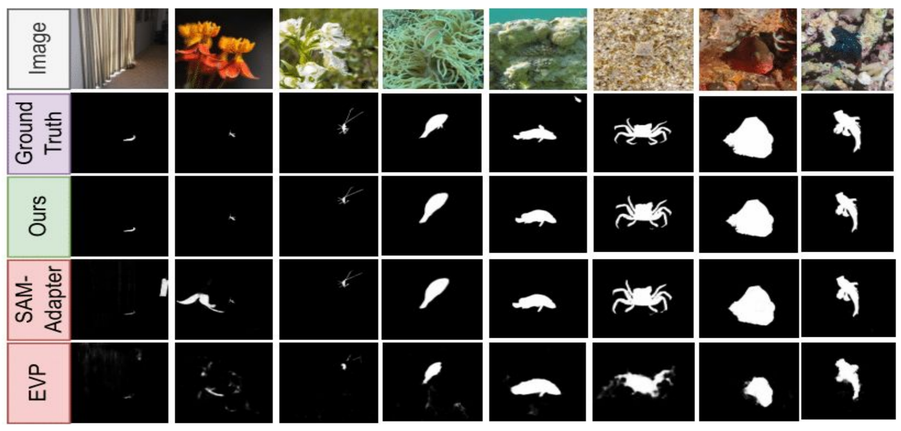
$\mathbb{C}$-Adapters

$$Emb_{comp} = f_{comp}(I_{HF})$$

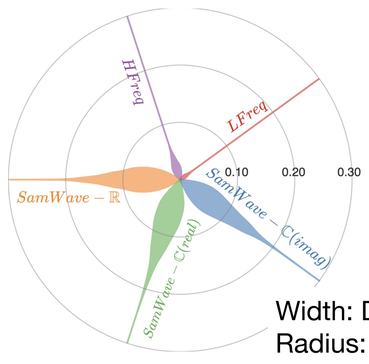
Complex-valued

EXPERIMENTS & RESULTS

Method	CHAMELEON				CAMO				COD10K			
	$\mathcal{S}_m\uparrow$	$\mathcal{E}_\phi\uparrow$	$\mathcal{F}_\beta^w\uparrow$	$\mathcal{M}\downarrow$	$\mathcal{S}_m\uparrow$	$\mathcal{E}_\phi\uparrow$	$\mathcal{F}_\beta^w\uparrow$	$\mathcal{M}\downarrow$	$\mathcal{S}_m\uparrow$	$\mathcal{E}_\phi\uparrow$	$\mathcal{F}_\beta^w\uparrow$	$\mathcal{M}\downarrow$
SAM[23]	0.727	0.734	0. 639	0.081	0.684	0.687	0.606	0.132	0.783	0.798	0.701	0.050
SAM2[45]	0.359	0.375	0.115	0.357	0.350	0.411	0.079	0.311	0.429	0.505	0.115	0.218
SAM as Backbone												
Real-Valued Adapters												
Ours (daubechies (db))	0.922	0.947	0.856	0.023	0.865	0.903	0.790	0.057	0.900	0.936	0.824	0.021
Ours (coiflet (cf))	0.923	0.952	**0.890**	0.022	0.856	0.888	0.807	0.060	**0.903**	0.939	0.857	0.019
Ours (haar (ha))	0.923	0.954	0.866	0.023	0.868	0.905	0.798	0.056	0.899	0.934	0.819	0.022
Ours (symlet (sym))	0.915	0.947	0.845	0.027	0.846	0.880	0.777	0.067	0.897	0.938	0.822	0.021
Complex-Valued Adapters												
Ours (Symmetric-b (sy-b))	**0.931**	0.961	0.883	0.020	0.856	0.896	0.797	0.058	0.902	0.940	0.843	0.019
Ours (Symmetric-a (sy-a))	**0.928**	0.957	0.881	0.019	0.858	0.893	0.800	0.057	0.902	0.939	0.841	0.019
SAM2 as Backbone												
Real-Valued Adapters												
Ours (daubechies (db))	0.912	0.952	0.880	0.018	**0.870**	**0.927**	0.828	**0.045**	**0.905**	**0.956**	0.861	**0.016**
Ours (coiflet (cf))	0.917	**0.963**	0.883	**0.017**	0.866	0.923	0.824	0.048	**0.905**	**0.957**	**0.862**	**0.016**
Ours (haar (ha))	0.918	0.962	**0.889**	**0.017**	0.869	0.926	**0.829**	**0.047**	0.900	0.952	**0.865**	**0.017**
Ours (symlet (sym))	0.917	0.961	0.887	**0.017**	**0.871**	**0.929**	**0.831**	**0.045**	**0.905**	0.954	0.859	**0.016**
Complex-Valued Adapters												
Ours (Symmetric-b (sy-b))	0.917	**0.966**	**0.890**	**0.016**	0.864	0.921	0.824	0.049	0.902	0.955	0.858	**0.017**
Ours (Symmetric-a (sy-a))	0.917	0.960	0.881	0.019	0.864	0.921	0.826	0.048	**0.903**	0.955	0.856	**0.016**



By Examining the weights of adapters, we found that weight values are evenly disturbed in contrast to previous method.



Width: Density
Radius: Weight

Effect of each high-frequency component

Method				CHAMELEON				CAMO				COD10K			
				$\mathcal{S}_m\uparrow$	$\mathcal{E}_\phi\uparrow$	$\mathcal{F}_\beta^w\uparrow$	$\mathcal{M}\downarrow$	$\mathcal{S}_m\uparrow$	$\mathcal{E}_\phi\uparrow$	$\mathcal{F}_\beta^w\uparrow$	$\mathcal{M}\downarrow$	$\mathcal{S}_m\uparrow$	$\mathcal{E}_\phi\uparrow$	$\mathcal{F}_\beta^w\uparrow$	$\mathcal{M}\downarrow$
SAM Finetuned				0.796	0.802	0.676	0.062	0.750	0.756	0.639	0.105	0.789	0.817	0.596	0.049
LL	*HL*	*LH*	*HH*	Evaluating each component											
✔	✗	✗	✗	0.907	0.937	0.842	0.029	**0.856**	0.884	0.785	0.062	0.891	0.929	0.817	0.024
✗	✔	✗	✗	0.895	0.927	0.834	0.030	0.851	0.882	0.781	0.063	0.887	0.924	0.816	0.024
✗	✗	✔	✗	0.896	0.915	0.838	0.030	0.854	0.887	0.790	0.063	0.891	0.928	0.838	0.023
✗	✗	✗	✔	0.908	0.935	0.848	0.028	0.853	0.883	0.783	0.064	0.893	0.928	0.818	0.023
✗	✔	✔	✔	**0.923**	**0.952**	**0.890**	**0.022**	**0.856**	**0.888**	**0.807**	**0.060**	**0.903**	**0.939**	**0.857**	**0.019**