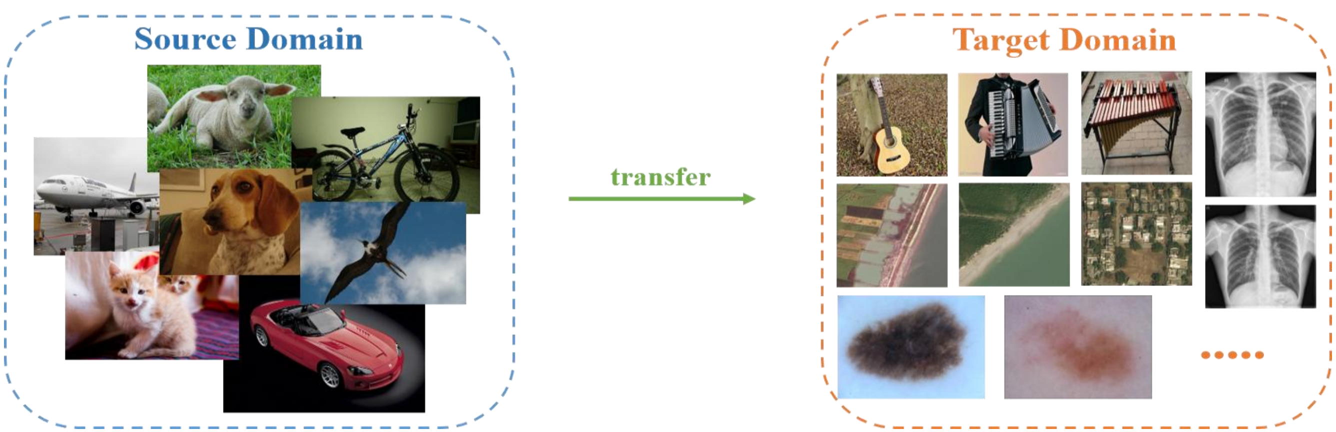
Lightweight Frequency Masker for Cross-Domain Few-Shot Semantic Segmentation

Jintao Tong, Yixiong Zou , Yuhua Li and Ruixuan Li
School of Computer Science and Technology, Huazhong University of Science and Technology
{jintaotong, yixiongz, idcliyuhua, rxli}@hust.edu.cn

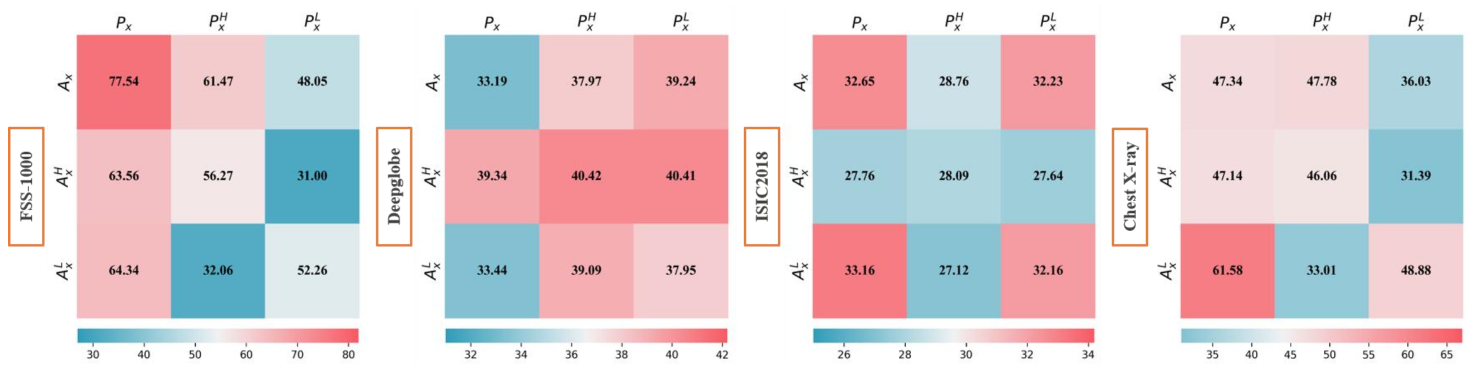
Cross-Domain Few-Shot Semantic Segmentation (CD-FSS)

- Setting
  - Trained on the source domain
  - Tested on the target domain
  - Different data distributions between source and target domain
- Task
  - Segment unseen classes from target domain
- Difficulty
  - Limited target data
  - Huge domain gap between source data and target data



Motivation

- Study the Domain Shift Problem
  - From the perspective of the frequency domain
- Intriguing Phenomenon
  - Simply filtering different frequency components for target domains can lead to a significant performance improvement



P: *Phase* A: *Amplitude* H: *High Frequency* L: *Low Frequency*

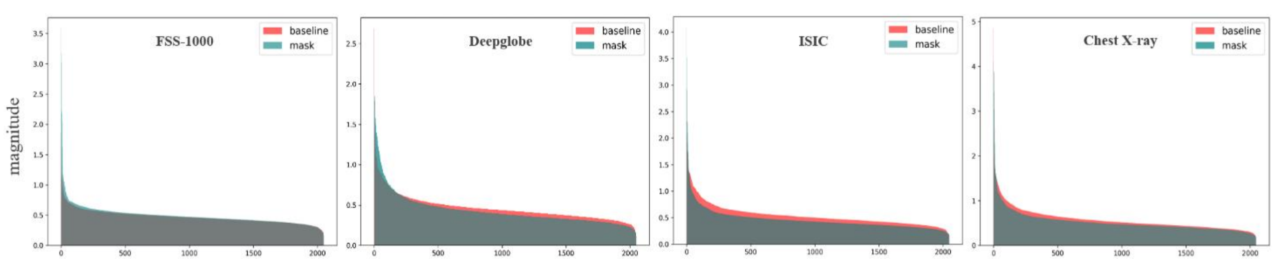
baseline $(A_x,\ P_x)$: *without filtering out any frequency components*

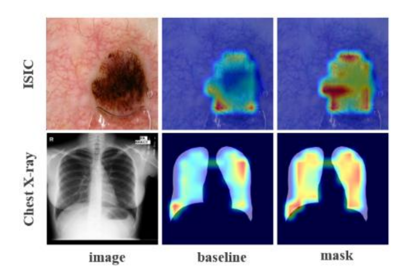
Delve into this phenomenon

- Enhanced Performance Stem from Reduced Inter-Channel Correlation
  - Different feature channels can represent distinct patterns
  - Performance improves as inter-channel mutual information (MI) decreases

Dataset	FSS-1000			Deepglobe			ISIC			Chest X-ray		
	baseline	best	worst	baseline	best	worst	baseline	best	worst	baseline	best	worst
1-shot MIoU	77.54	64.34↓	48.054↓	33.19	40.42↑	33.44↑	32.65	33.16↑	27.124↓	47.34	61.58↑	31.394↓
support MI	1.3736	1.3791↑	1.8767↑	1.3679	1.35024↓	1.35584↓	1.3789	1.36974↓	1.3951↑	1.3952	1.39304↓	1.4315↑
query MI	1.3739	1.3805↑	1.8201↑	1.3667	1.34354↓	1.35984↓	1.3792	1.36944↓	1.3890↑	1.3921	1.38774↓	1.4368↑

- Why Lower Inter-Channel Correlation is Better?
  - Improve cross-domain generalization
    - More uniform Mean Magnitude of Channels (MMC) curve
    - Handle channel bias problem
    - More independent and diverse semantic patterns
  - Enlarge activation regions for segmentation
    - Better detect the entire object





Feature Disentanglement in the Frequency Domain

- Mathematical Derivation
  - Prove the correlation between phase differences and channel correlation
    - Fourier Transform (FT)

$$F(u,v) = \frac{1}{wh}\sum_{x=0}^{w-1}\sum_{y=0}^{h-1} f(x,y)e^{-i2\pi\left(\frac{ux}{w}+\frac{vy}{h}\right)}$$

$$F = \alpha\cos(\rho) + i\alpha\sin(\rho) = \alpha \cdot e^{i\rho}$$

$$|F| = \sqrt{\alpha^2(\cos^2(\rho)+\sin^2(\rho))} = \sqrt{\alpha^2} = \alpha$$

    - The correlation coefficient formula in the frequency domain:

$$r = \sum_{m=0}^{h-1}\sum_{n=0}^{w-1} \frac{F_1(m,n)F_2^*(m,n)}{\sqrt{|F_1(m,n)|^2|F_2(m,n)|^2}} = \sum_{m=0}^{h-1}\sum_{n=0}^{w-1} r(m,n)$$

    - The complex conjugate

$$F_2^*(m,n) = \alpha_2\cos(\rho_2) - i\alpha_2\sin(\rho_2) = \alpha_2 e^{-i\rho_2}$$

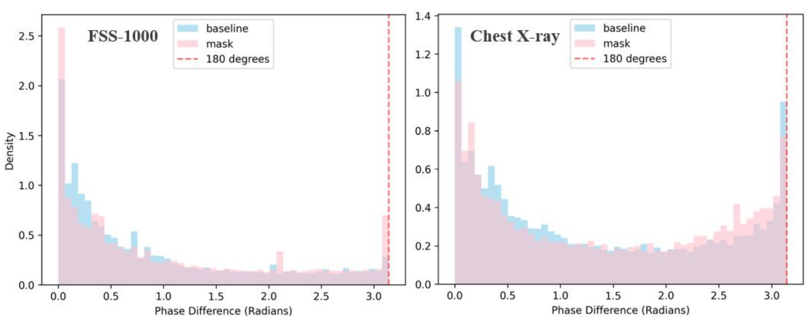
$$F_1(m,n)F_2^*(m,n) = \alpha_1\alpha_2 e^{i(\rho_1-\rho_2)}$$

$$|F_1(m,n)|^2|F_2(m,n)|^2 = \alpha_1^2\alpha_2^2$$

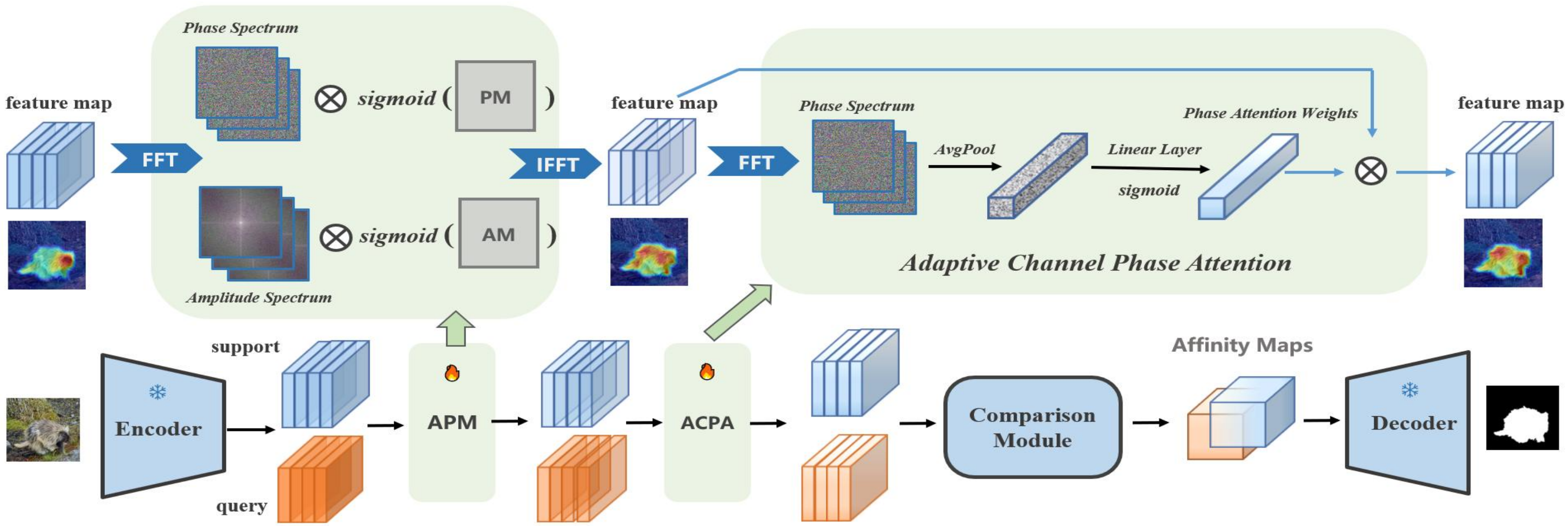
    - Derivation result:

$$r(m,n) = \frac{\alpha_1\alpha_2 e^{i(\rho_1-\rho_2)}}{\sqrt{\alpha_1^2\alpha_2^2}} = e^{i(\rho_1-\rho_2)},\quad \rho_1-\rho_2 = \Delta\rho \in [0,\pi]$$

    - Experiments for Derivation



Method

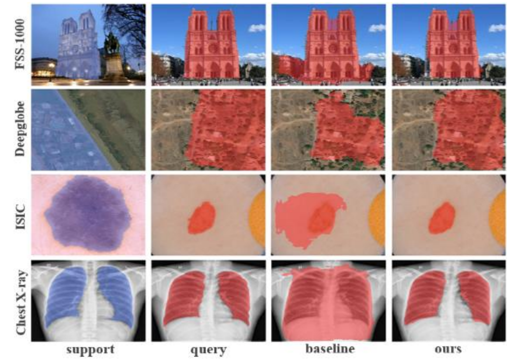


- Lightweight Frequency Masker
- Amplitude-Phase Masker (APM)
  - Reduce inter-channel correlation
  - Accomplish feature disentanglement
  - Obtain more independent semantic representations
- Adaptive Channel Phase Attention (ACPA)
  - Leverage phase invariance
  - Align the support and query feature spaces

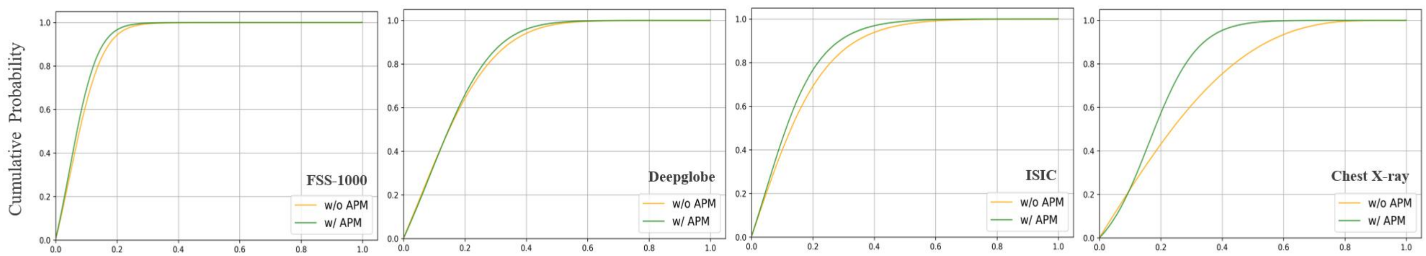
Experiments

- State-of-the-Art Performance

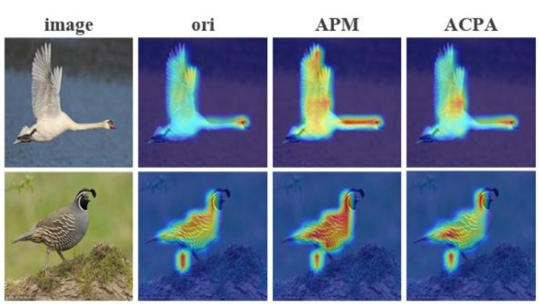
Method	FSS-1000		Deepglobe		ISIC		Chest X-ray		Average	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
PGNet [45]	62.42	62.74	10.73	12.36	21.86	21.25	33.95	27.96	32.24	31.08
PANet [41]	69.15	71.68	36.55	**45.43**	25.29	33.99	57.75	69.31	47.19	55.10
CaNet [46]	70.67	72.03	22.32	23.07	25.16	28.22	28.35	28.62	36.63	37.99
RPMMs [43]	65.12	67.06	12.99	13.47	18.02	20.04	30.11	30.82	31.56	32.85
PFENet [37]	70.87	70.52	16.88	18.01	23.50	23.83	27.22	27.57	34.62	34.98
RePRI [5]	70.96	74.23	25.03	27.41	23.27	26.23	65.08	65.48	46.09	48.34
HSNet [31]	77.53	80.99	29.65	35.08	31.20	35.10	51.88	54.36	47.57	51.38
HSNet* [31]	77.54	80.21	33.19	36.46	32.65	35.09	47.34	48.63	47.68	50.10
PATNet [24]	78.59	81.23	37.89	42.97	41.16	**53.58**	66.61	70.20	56.06	61.99
Ours (APM-S)	78.25	80.29	40.77	44.85	41.48	49.39	75.22	76.89	58.93	62.86
Ours (APM-M)	**79.29**	**81.83**	**40.86**	44.92	**41.71**	51.16	**78.25**	**82.81**	**60.03**	**65.18**



- APM: Feature Disentanglement via Frequency Operations



- ACPA: Aligning Task-Relevant Features and Feature Spaces



APM-S	APM-M	ACPA	FSS	Deepglobe	ISIC	Chest
			0.3591	0.2691	0.2494	0.5848
✓			0.3481	0.2678	0.2433	0.5025
✓		✓	0.2907	0.2676	0.2032	0.3628
	✓		0.3293	0.2675	0.2310	0.4635
	✓	✓	**0.2883**	**0.2674**	**0.1811**	**0.3986**

Thanks!