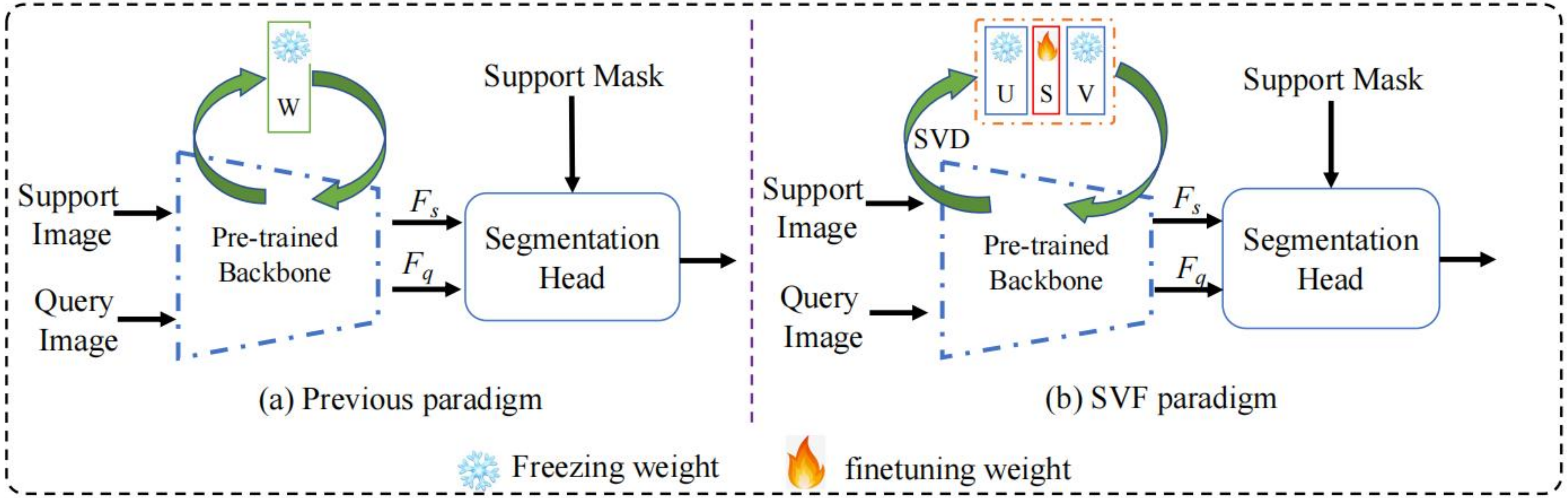
01 Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning (NeurIPS 2022)

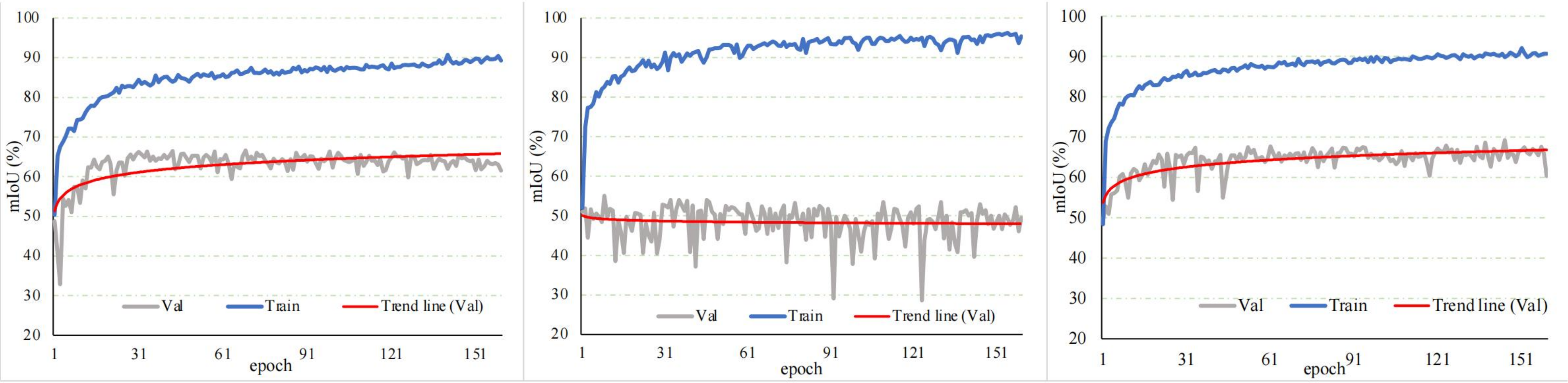
Yanpeng Sun[1*], Qiang Chen[2*], Xiangyu He[3*], Jian Wang[2], Haocheng Feng[2]
Junyu Han[2], Errui Ding[2], Jian Cheng[3], Zechao Li[1†], Jingdong Wang[2]
[1]School of Computer Science and Engineering, Nanjing University of Science and Technology
[2]Baidu VIS
[3]NLPR, Institute of Automation, Chinese Academy of Sciences



直接采用图像分类的预训练参数来进行小样本分割是否是唯一的选择？

01 Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning

- **Fixed** vs **Fully finetune** vs **SVF finetune**
  - 传统fine-tune方法使backbone更加适应training set中的类，从而影响了模型泛化到test set的能力
  - SVF：仅fine-tune backbone中的一小部分参数，使得backbone更加适应当前任务，并保持backbone的泛化能力



(a) PFENet (b) PFENet + fully finetune (b) PFENet + SVF

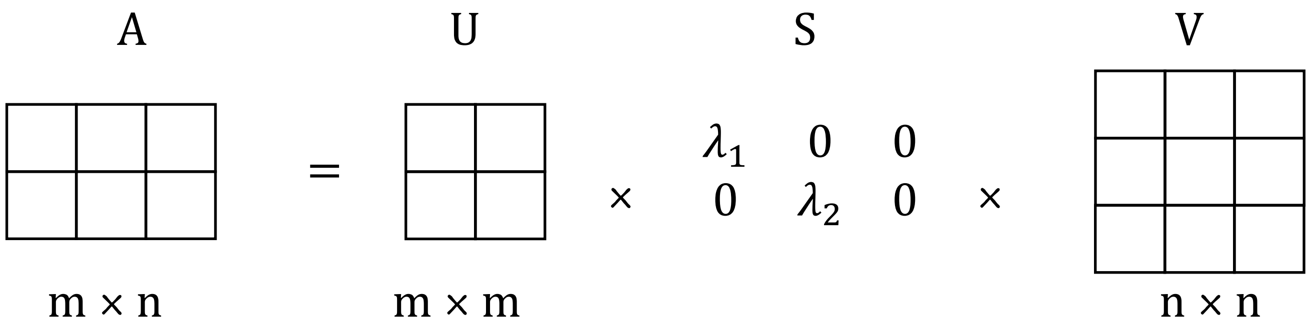
01 Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning

- **SVD 分解 （奇异值分解）**
  - 任何m × n 大小矩阵可进行如下分解：

$$A_{\mathrm{m\times n}} = U_{\mathrm{m\times m}}\, S\, V_{\mathrm{n\times n}}$$

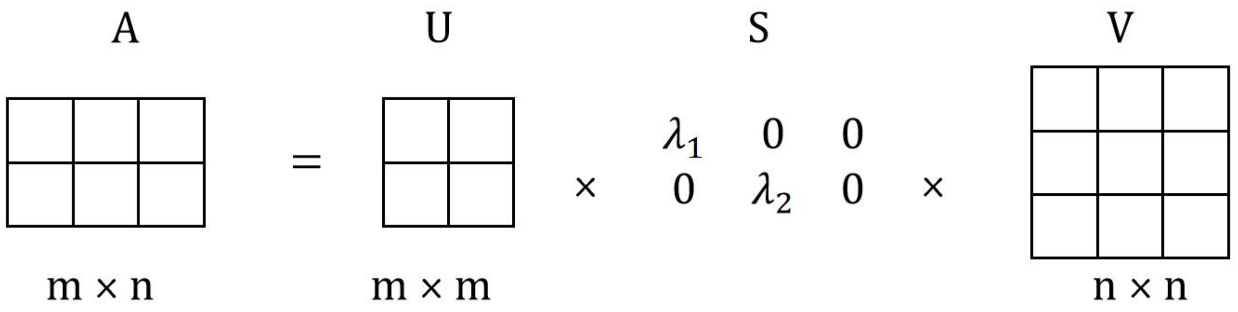
U, V 为正交矩阵，$UU^T = VV^T = I$(单位矩阵)

奇异矩阵 S，非对角线为0



01 Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning

- SVD 分解 数学举例



用一个简单的例子来详细化矩阵的奇异值分解过程。假如一个矩阵A为：

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$$

先计算A^TA和AA^T

$$A^TA = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$AA^T = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

然后求解A^TA的特征值及对应的特征向量：

$$\lambda_1 = 3; v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}; \lambda_2 = 1; v_2 = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

同理求解AA^T的特征值及对应的特征向量：

$$\lambda_1 = 3; u_1 = \begin{pmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}; \lambda_2 = 1; u_2 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix}; \lambda_3 = 0; u_3 = \begin{pmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}$$

通过$\sigma_i = \sqrt{\lambda_i}$求解奇异值为$\sqrt{3}$和1

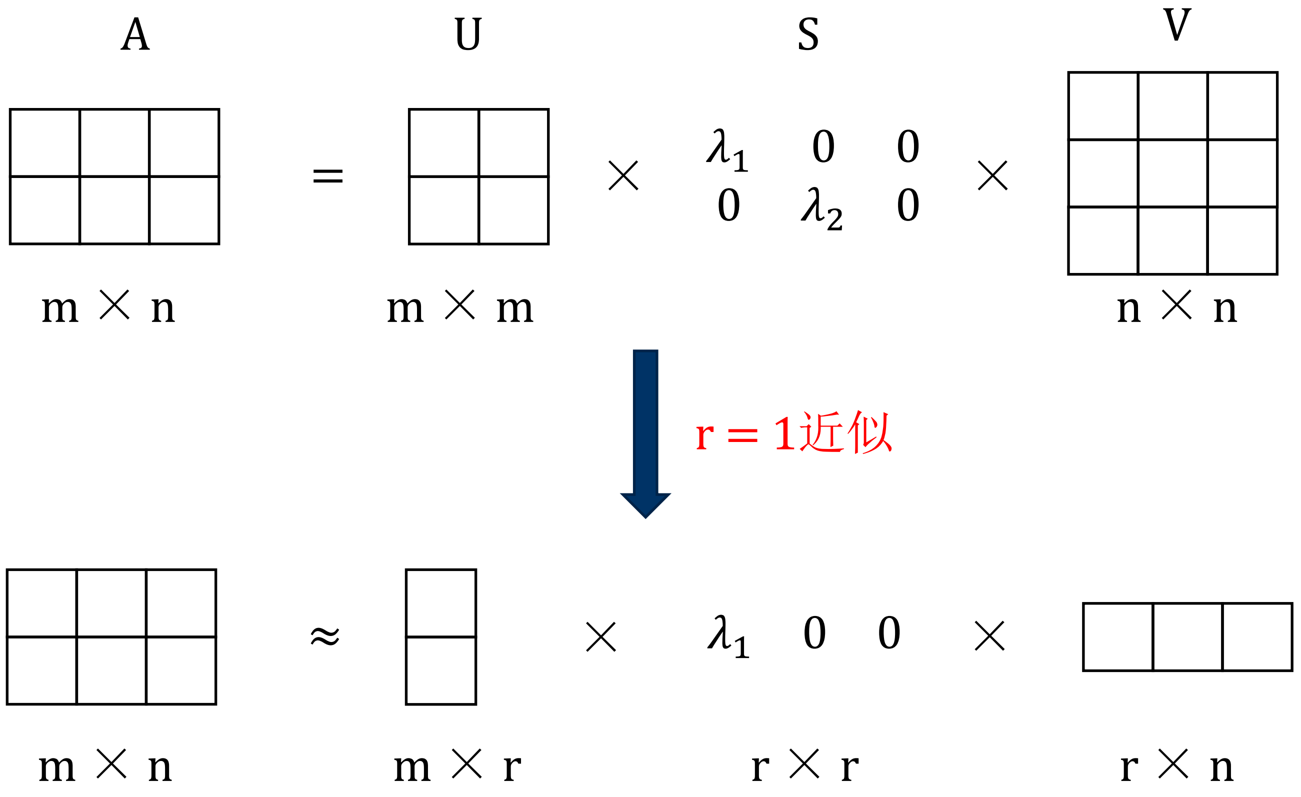
最终矩阵A的奇异值分解为：

$$A = U\Sigma V^T = \begin{pmatrix} 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

01 Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning

- SVD 分解 矩阵压缩
  - 压缩较小奇异值：

$$A_{m\times n} = U_{m\times m}\, S\, V_{n\times n} \Longrightarrow A_{m\times n} \approx U_{m\times r}\, S_{r\times r}\, V_{r\times n}$$



01 Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning

- **SVD 分解** 模型压缩
  - 将原来大的权重矩阵分解为多个小的矩阵，用低秩矩阵近似原有权重矩阵

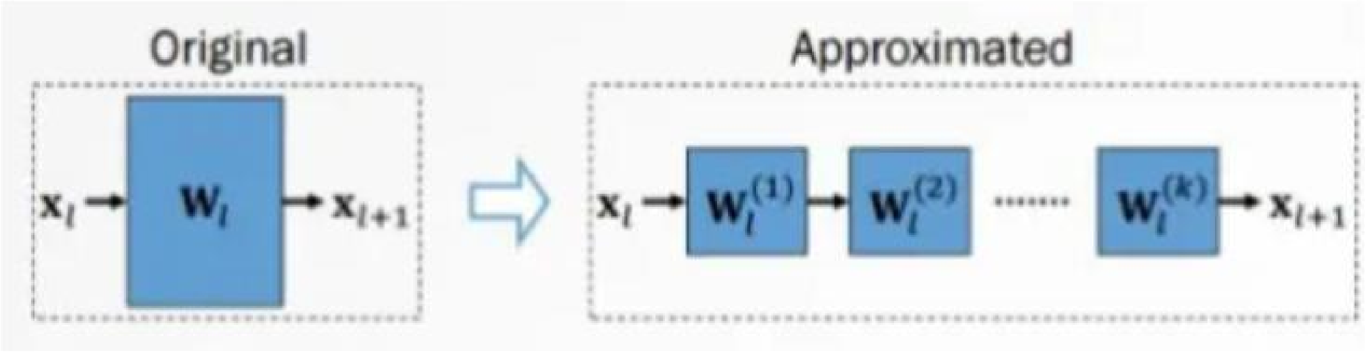
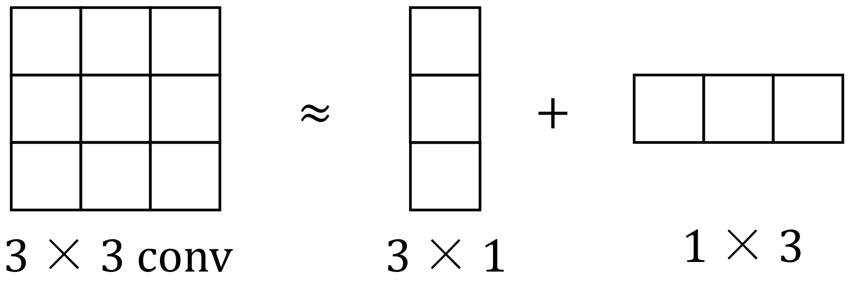


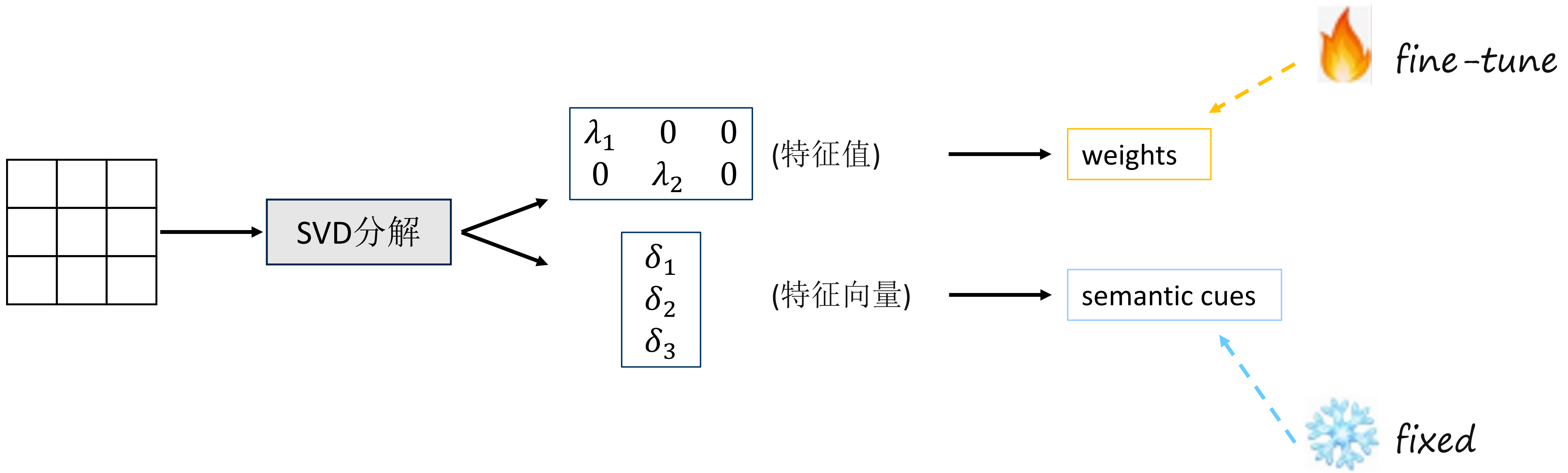
图1：低秩分解 Low Rank Expansion

e.g.



01 Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning

- **SVD 分解** **Few shot**
  - **核心思想**--不改变pre-train中的语义线索 (semantic cues)，仅改变不同语义线索的权重 (weights)



01 Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning

- **SVD 分解** **Few shot**

$$\mathbf{W} \in \mathbb{R}^{C_o \times C_i \times K \times K} \longrightarrow \mathbf{W}' \in \mathbb{R}^{C_o \times C_i K^2}$$

$$\mathbf{W}' = \mathbf{USV}^T, \text{ where } \boxed{\mathbf{U} \in \mathbb{R}^{C_o \times R}}, \boxed{\mathbf{S} \in \mathbb{R}^{R \times R}}, \text{ and } \boxed{\mathbf{V}^T \in \mathbb{R}^{R \times C_i K^2}}$$

A K × K conv layer corresponds to three successive layers:

- a R × Ci × K × K convolution layer,
- a scaling layer,
- a Co × R × 1 × 1 convolution layer

Algorithm 1 Pseudocode of SVF in Python style

```
# Input: Conv2d with weight matrix W, Input feature X
# Output: Output feature Y

# Previous 3x3 Conv :
# Y = F.Conv2d(W, X, kernel=(3,3))

# SVF :
U, S, V = svd(W) # decompose weights by SVD

U.requires_grad = False # freeze Conv_U
V.requires_grad = False # freeze Conv_V

Y = F.Conv2d(V, X, kernel=(3,3)) # a new 3x3 conv
Y = Y.mul(S) # reconstruct a new affine layer
Y = F.Conv2d(U, Y, kernel=(1,1)) # a new 1x1 conv
```

01 Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning

- **Discussion on Why SVF Works**

预先训练好的主干中包含的语义线索可能与支持图像中显示的对象无关，这给在FSS中分割新的类对象带来了意想不到的障碍。

Figure 6: Statistics chart about the changes of initial Top-30 largest singular values of the last 1 × 1 and 3 × 3 convolution layer in layer3 after fine-tuning.

对应背景响应的奇异值减小，对应前景响应的奇异值增加

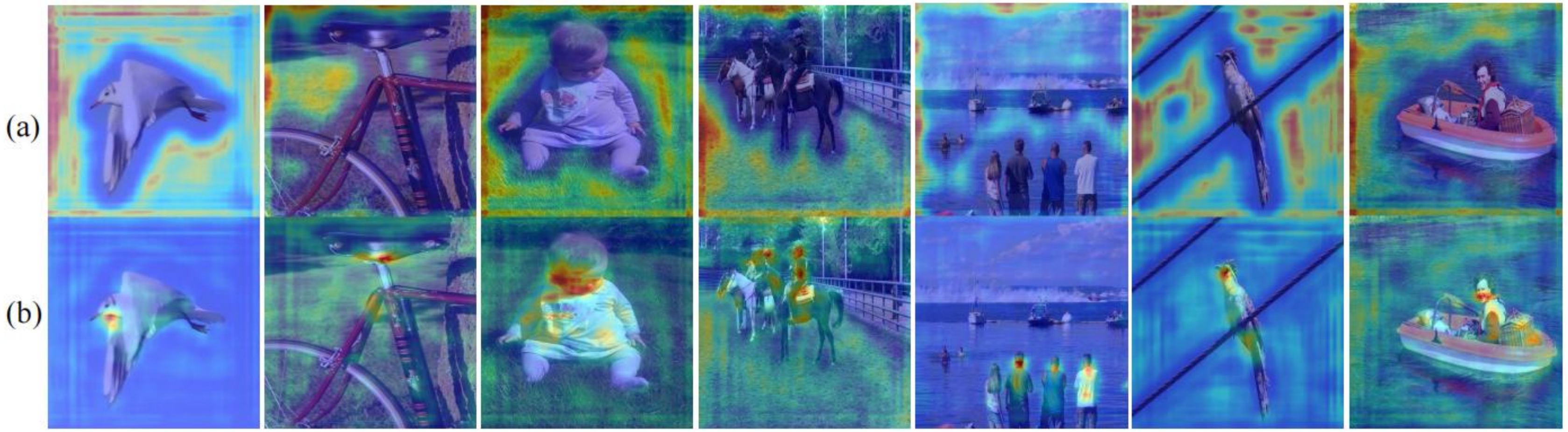


Figure 4: The visualization of segmentation cues with the largest variation in singular values from the last 3 × 3 convolution in layer 3. (a) represents segmentation clues of subspace **U** with the largest singular value reduction, (b) represents segmentation clues of subspace **U** with the largest singular value growth.

01 Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning

Table 1: Performance on Pascal-5^i [33] in terms of mIoU for 1-shot and 5-shot segmentation. The best mean results are show in **bold**. † indicates that images from training set containing the novel class on test set were removed.

Method	backbone	1-shot					5-shot				
		Fold-0	Fold-1	Fold-2	Fold-3	Mean	Fold-0	Fold-1	Fold-2	Fold-3	Mean
baseline†	VGG16	57.48	66.72	62.66	53.72	60.15	62.98	70.57	68.62	59.60	65.44
baseline†+SVF	VGG16	63.07	68.40	65.81	54.28	**62.89**(+2.74)	68.52	72.15	69.08	63.59	**68.34**(+2.90)
PFENet† [39]	VGG16	61.91	70.34	63.77	57.38	63.35	67.73	72.82	69.31	67.59	69.36
PFENet†+SVF	VGG16	63.43	71.40	64.18	58.30	**64.33**(+0.98)	69.11	73.67	69.13	67.30	**69.80**(+0.44)
BAM† [16]	VGG16	63.18	70.77	66.14	57.53	64.41	67.36	73.05	70.61	64.00	68.76
BAM†+SVF	VGG16	64.09	71.07	66.79	57.54	**64.87**(+0.46)	67.75	74.11	70.99	63.57	**69.11**(+0.35)
baseline†	ResNet50	65.60	70.28	64.12	60.27	65.07	69.89	74.16	67.87	65.73	69.41
baseline†+SVF	ResNet50	67.42	71.57	67.99	61.57	**67.14**(+2.07)	70.37	75.06	71.08	69.16	**71.42**(+2.01)
baseline	ResNet50	66.36	69.22	57.64	58.73	62.99	70.75	72.92	58.86	65.56	67.02
baseline + SVF	ResNet50	66.88	70.84	62.33	60.63	**65.17**(+2.18)	71.49	74.04	59.38	67.43	**68.09**(+1.07)
PFENet† [39]	ResNet50	66.61	72.55	65.33	60.91	66.35	70.93	75.32	69.60	68.96	71.20
PFENet†+SVF	ResNet50	69.27	73.55	67.49	62.30	**68.15**(+1.80)	71.82	74.92	70.97	69.58	**71.82**(+0.62)
PFENet	ResNet50	67.06	71.61	55.21	59.46	63.34	72.11	73.67	61.61	67.50	68.72
PFENet + SVF	ResNet50	68.31	71.99	56.25	61.82	**64.59**(+1.25)	72.09	73.99	63.58	70.03	**69.92**(+1.20)
BAM† [16]	ResNet50	68.97	73.59	67.55	61.13	67.81	70.59	75.05	70.79	67.20	70.91
BAM†+SVF	ResNet50	69.38	74.51	68.80	63.09	**68.95**(+1.14)	72.05	76.17	71.97	68.91	**72.28**(+1.37)
BAM	ResNet50	68.37	72.05	57.55	60.38	64.59	70.72	74.21	63.58	66.18	68.67
BAM + SVF	ResNet50	68.17	72.86	57.77	62.04	**65.21**(+0.62)	72.30	74.43	65.16	69.43	**70.33**(+1.66)

01 Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning

Table 2: Performance on COCO-20^i [27] in terms of mIoU for 1-shot and 5-shot segmentation. The best mean results are show in **bold**. † indicates that images from training set containing the novel class on test set were removed.

Method	Backbone	1-shot					5-shot				
		Fold-0	Fold-1	Fold-2	Fold-3	Mean	Fold-0	Fold-1	Fold-2	Fold-3	Mean
baseline†	VGG-16	37.56	37.73	38.78	35.66	37.43	43.07	49.42	44.38	46.38	45.81
baseline†+SVF	VGG-16	39.32	39.64	38.63	35.45	**38.26**$_{(+0.83)}$	46.48	50.72	45.79	45.63	**47.16**$_{(+1.35)}$
PFENet [39]	VGG-16	35.40	38.10	36.80	34.70	36.30	38.20	42.50	41.80	38.90	40.40
PFENet† [39]	VGG-16	41.03	44.22	43.74	38.90	41.97	48.66	48.26	45.49	51.02	48.36
PFENet†+SVF	VGG-16	42.68	44.90	42.60	38.79	**42.24**$_{(+0.27)}$	49.02	53.71	47.59	47.63	**49.49**$_{(+1.13)}$
BAM† [16]	VGG-16	38.96	47.04	46.41	41.57	43.50	47.02	52.62	48.59	49.11	**49.34**
BAM†+SVF	VGG-16	40.21	46.62	46.23	41.97	**43.76**$_{(+0.26)}$	45.05	53.59	48.35	49.28	49.07$_{(-0.27)}$
baseline†	ResNet-50	38.91	46.07	42.67	39.71	41.84	50.35	56.78	49.61	50.96	51.93
baseline†+SVF	ResNet-50	44.22	46.38	42.65	41.65	**43.72**$_{(+1.88)}$	51.47	57.48	50.33	52.29	**52.89**$_{(+1.93)}$
PFENet† [39]	ResNet-50	44.93	50.32	44.68	44.26	46.05	52.29	59.34	51.50	53.53	54.17
PFENet†+SVF	ResNet-50	46.88	50.86	47.69	46.64	**48.02**$_{(+1.97)}$	52.72	58.14	52.52	54.15	**54.38**$_{(+0.21)}$
BAM† [16]	ResNet-50	43.41	50.59	47.49	43.42	46.23	49.26	54.20	51.63	49.55	51.16
BAM†+SVF	ResNet-50	46.87	53.80	48.43	44.78	**48.47**$_{(+2.24)}$	52.25	57.83	51.97	53.41	**53.87**$_{(+2.71)}$

01 Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning

Table 4: Ablation study of BN on Pascal-5^i under 1-shot setting. ✓represents fine-tuning this feature space. The best mean results are show in **bold**.

Method	BN	S	Mean
baseline†	-	-	65.07
	✓		63.12$_{(-1.95)}$
	✓	✓	64.20$_{(-0.87)}$
		✓	**67.14**$_{(+2.07)}$

Table 5: Comparative experiment with fine-tuning different layer of backbone on Pascal-5^i.

Method	layer	Mean
baseline†	-	65.07
+fully fine-tune	1, 2, 3, 4	60.90$_{(-4.17)}$
+ part fine-tune	2, 3, 4	61.15$_{(-3.92)}$
	3, 4	61.08$_{(-3.99)}$
	4	60.58$_{(-4.49)}$
+SVF	2, 3, 4	**67.14**$_{(+2.07)}$

Table 6: Comparative experiment with fine-tuning different convolutional layer of backbone on Pascal-5^i.

Method	layer	3×3	1×1	Mean
baseline†	-	-	-	65.07
+part fine-tune	2, 3, 4	✓	✓	61.15$_{(-3.92)}$
	2, 3, 4	✓		61.86$_{(-3.21)}$
	2, 3, 4		✓	61.62$_{(-3.45)}$
+SVF	2, 3, 4	-	-	67.14$_{(+2.07)}$

Table 7: Ablation study of SVF fine-tuning different subspace on Pascal-5^i.

Method	U	S	V	Mean
baseline†	✓			61.09
		✓		**67.14**
			✓	60.88
	✓	✓		61.57
		✓	✓	60.42
	✓		✓	60.02
	✓	✓	✓	61.24

Table 8: Ablation study of SVF fine-tuning different layer on Pascal-5^i. The best results are show in **bold**.

Method	layer	Mean
baseline† + SVF	4	66.21
	3, 4	**67.20**
	2, 3, 4	67.14
	1, 2, 3, 4	67.12

直接改变U或V子空间的特征分布会降低模型的泛化能力。