CVPR
VIRTUAL JUNE 14-19

The 2nd Learning from Imperfect Data (LID) Workshop

Revisiting Class Activation Mapping for Learning from Imperfect Data

Wonho Bae*, Junhyug Noh*, Jinhwan Seo, and Gunhee Kim

SEOUL NATIONAL UNIV.
VISION & LEARNING

Challenge Results

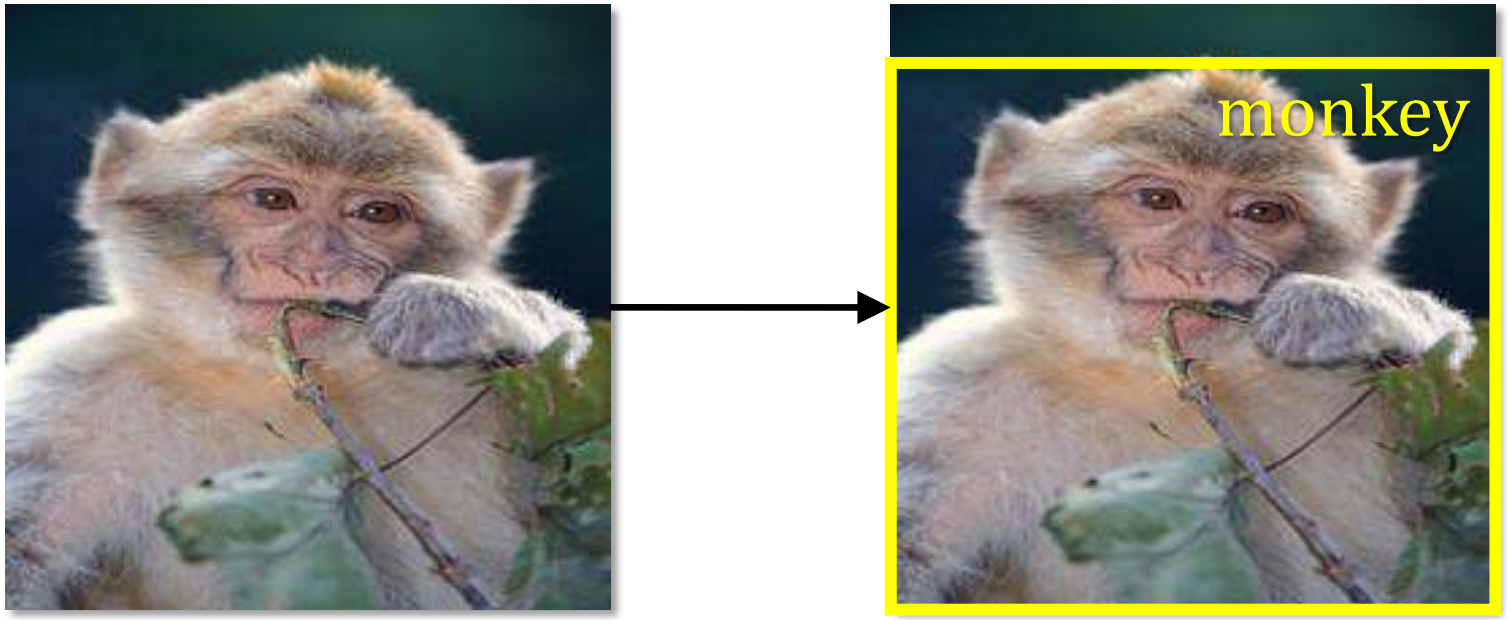
1st place

Track 3: Weakly Supervised Object Localization

2nd place

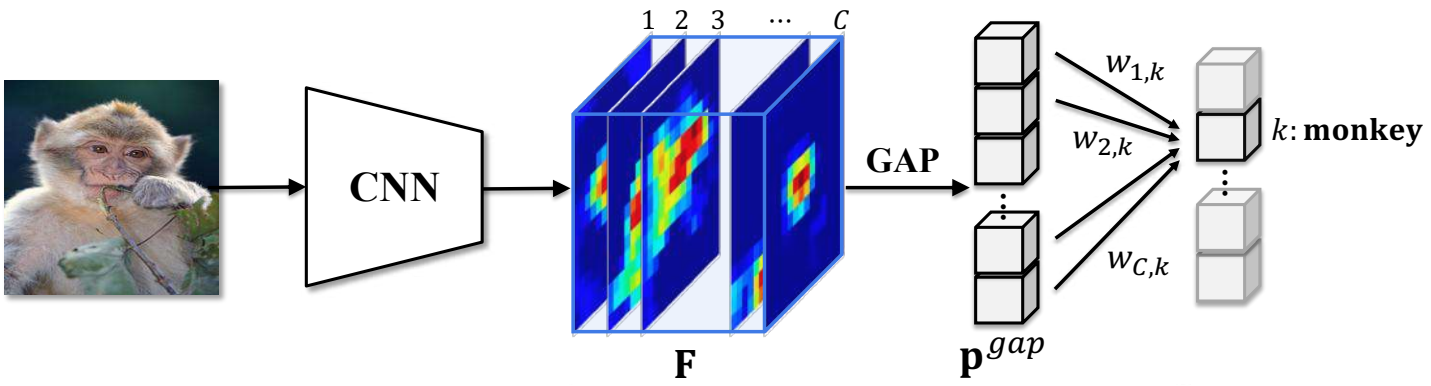
Track 1: Weakly Supervised Semantic Segmentation

Weakly-Supervised Object Localization

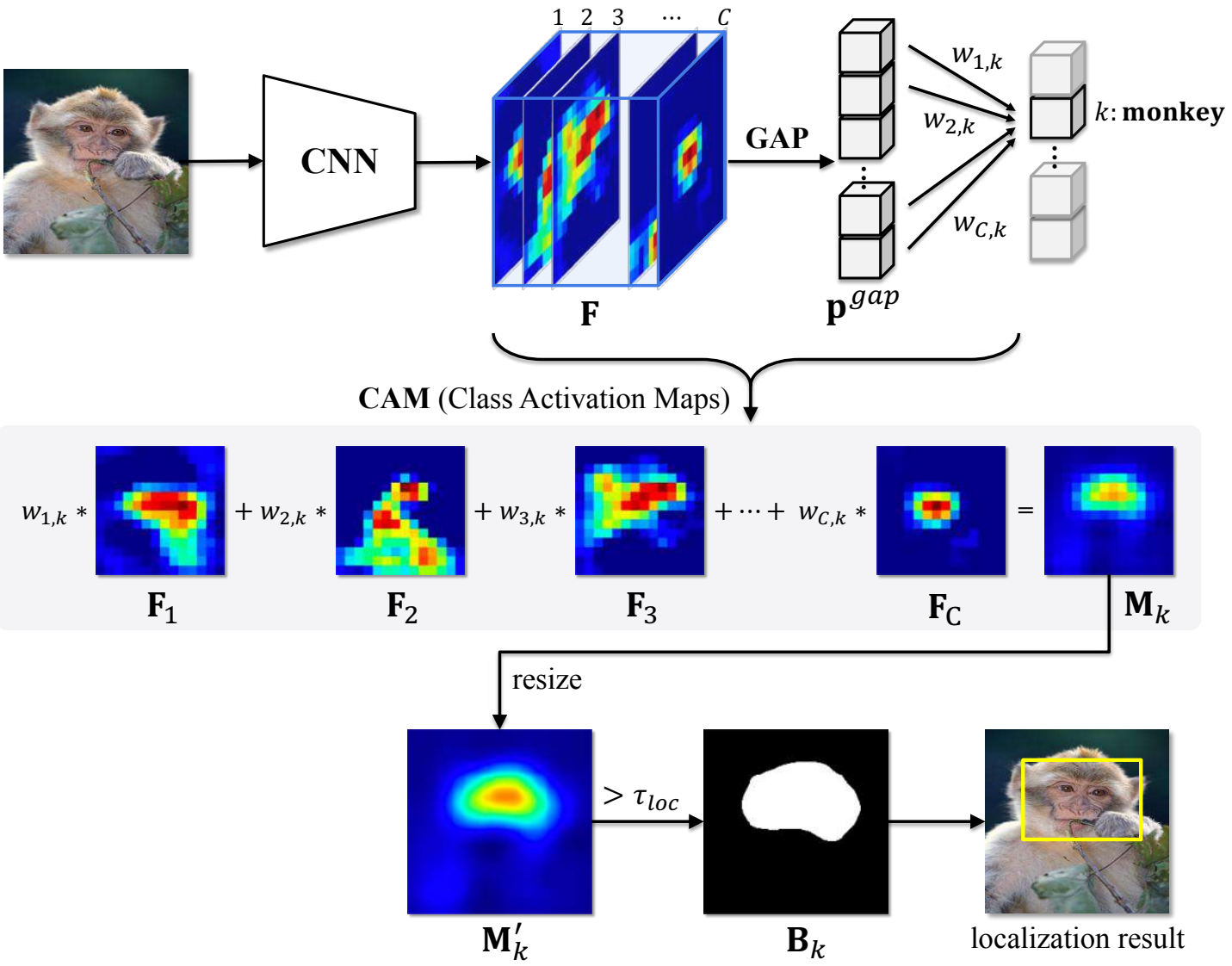


Input Output

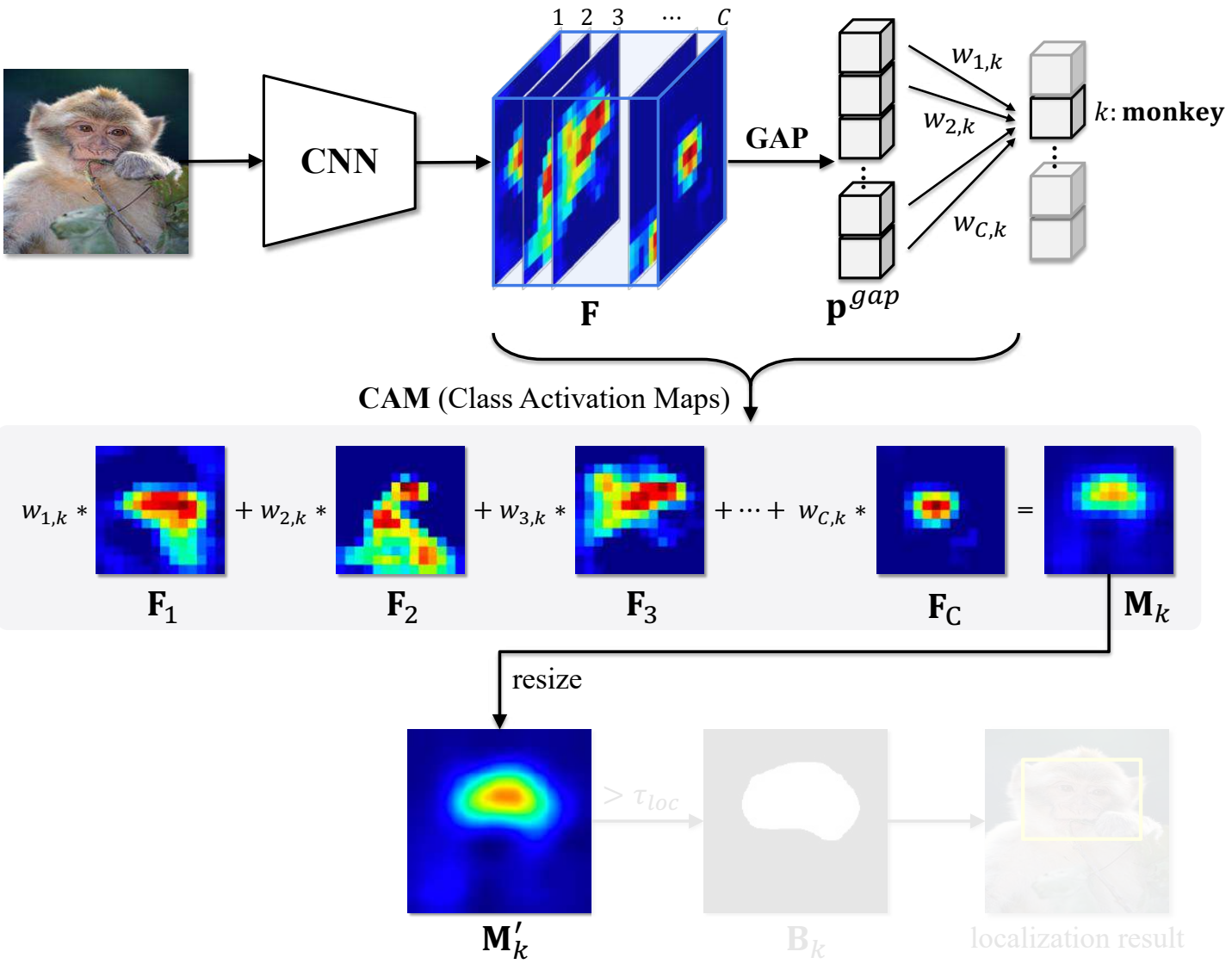
Class Activation Mapping (CAM)



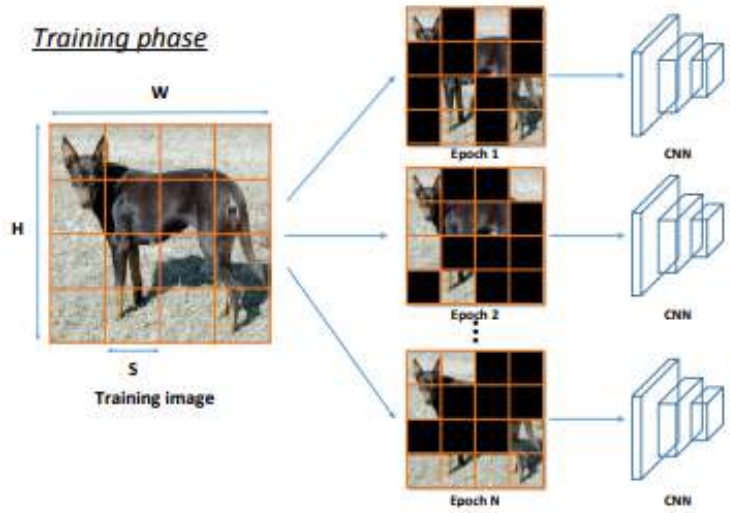
Class Activation Mapping (CAM)



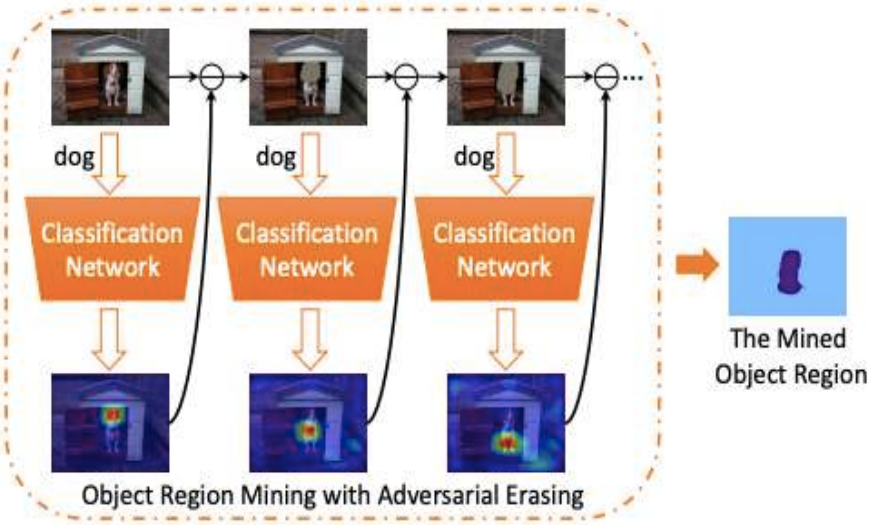
Class Activation Mapping (CAM) for Track 3



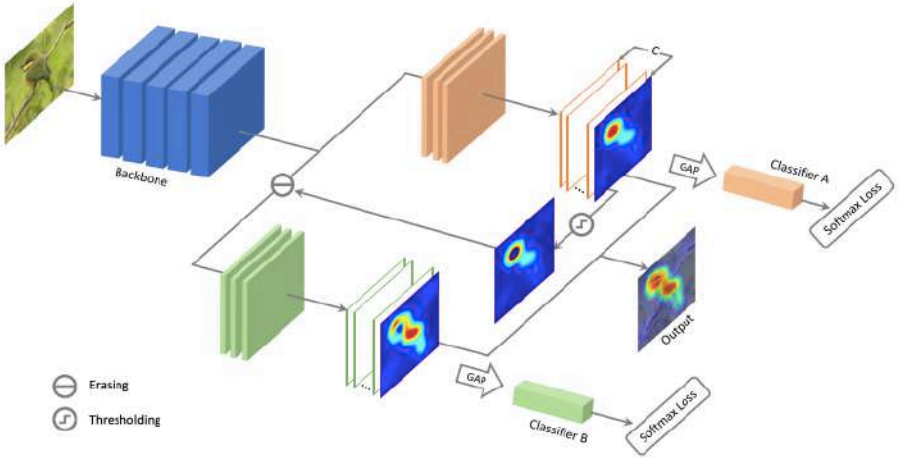
How to Grasp Whole Object Region?



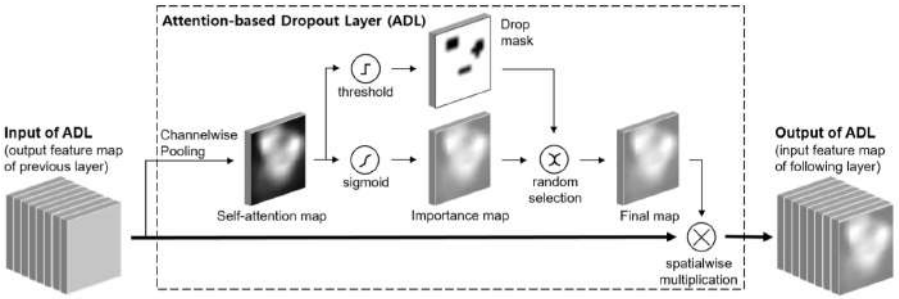
[HaS] Singh, et al. ICCV 2017



[AE] Wei, et al. CVPR 2017



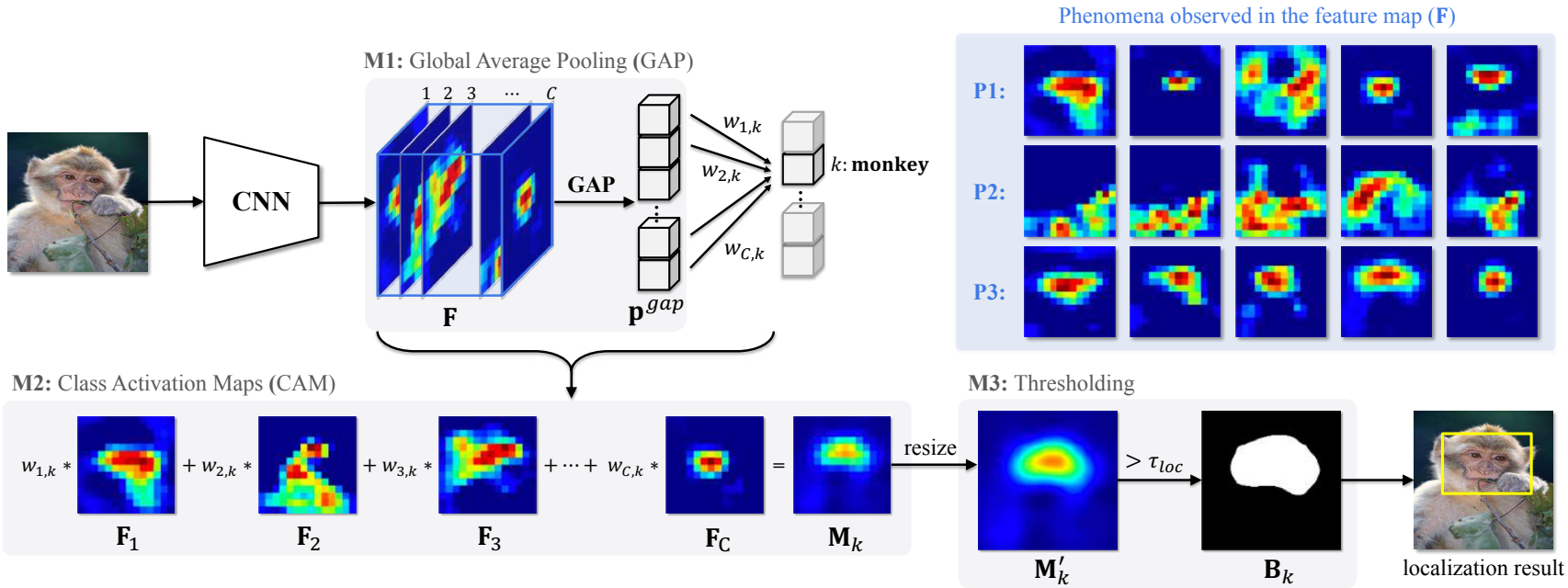
[ACoL] Zhang, et al. CVPR 2018



[ADL] Choe, et al. CVPR 2019

Our Approach

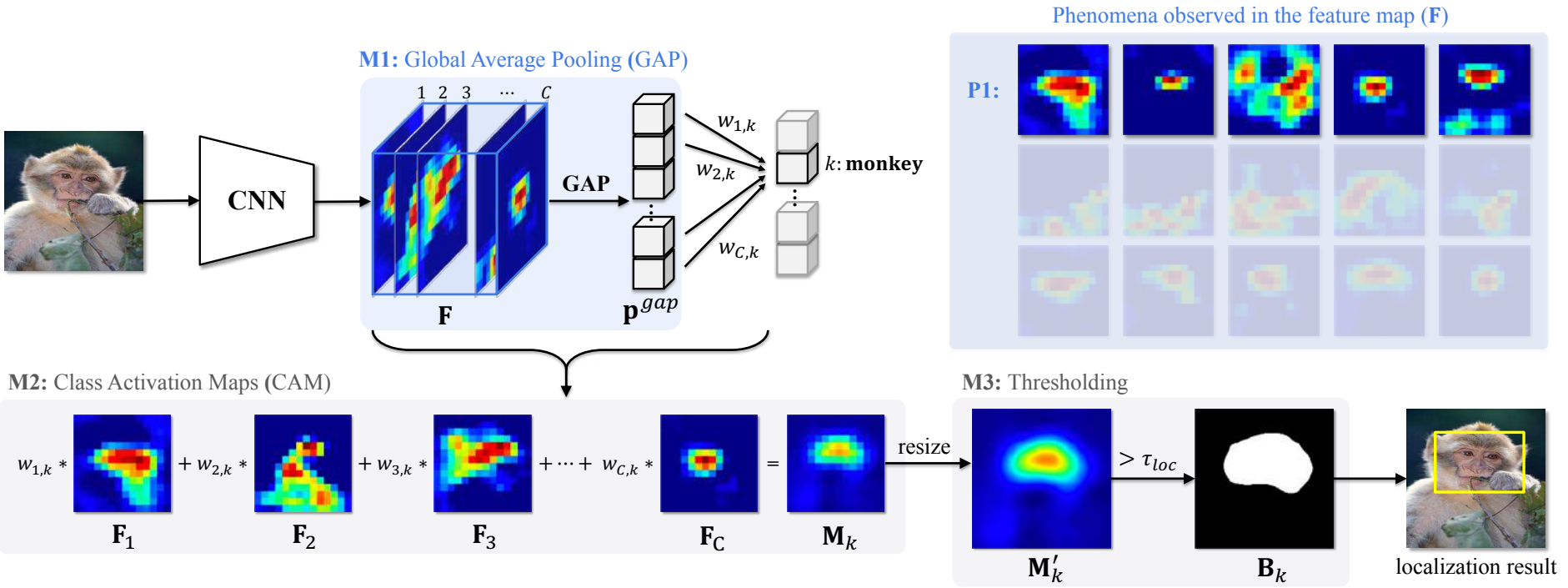
- **Motivation**
  - Information to capture the whole area of the object already exists in feature maps
- **Problem**
  - Three modules (M1–M3) of CAM do not take phenomena (P1–P3) into account
  - It results in the localization being limited to small discriminative regions of an object
- **Solution**
  - Correctly utilize the information by simply modifying the three modules



Our Approach (1) Thresholded Average Pooling

- **Problem**: Global Average Pooling **(GAP)** under **P1**

$$p_c^{\text{gap}} = \frac{1}{H \times W} \sum_{(h,w)} \mathbf{F}_c(h, w)$$

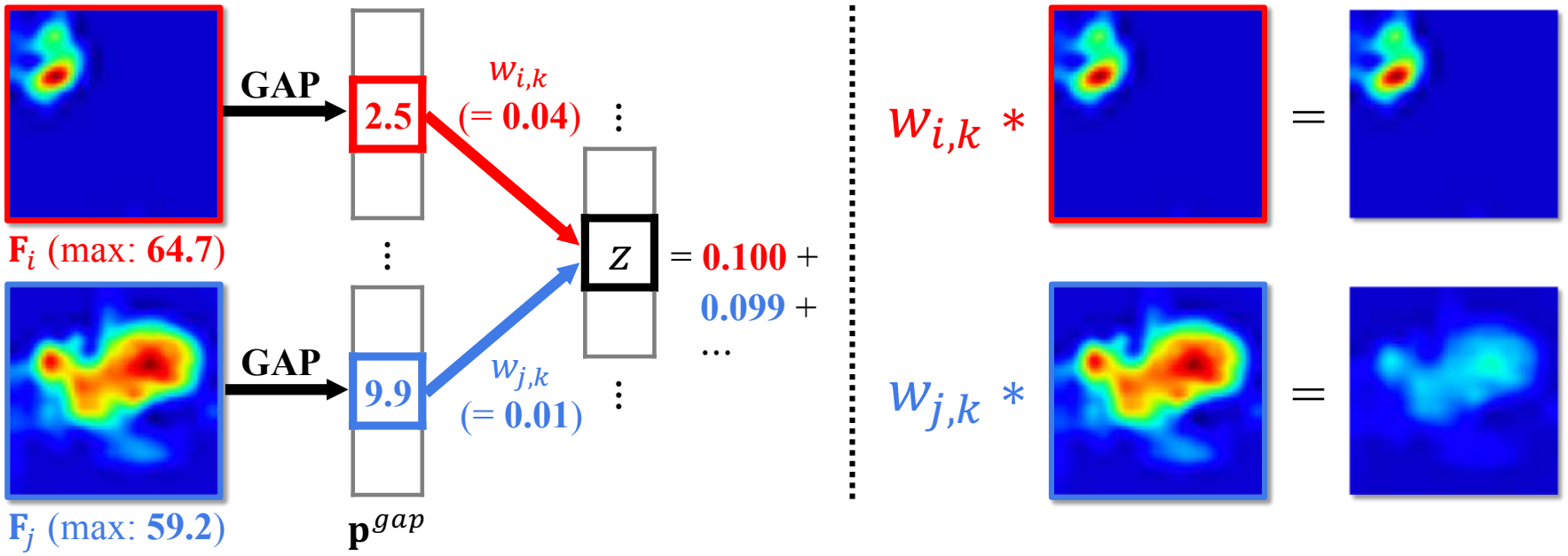


Our Approach (1) Thresholded Average Pooling

- **Problem**: Global Average Pooling **(GAP)** under **P1**

Our Approach (1) Thresholded Average Pooling

- **Problem**: Global Average Pooling **(GAP)** under **P1**



Classification phase

Localization phase

Our Approach (1) Thresholded Average Pooling

- **Problem**: Global Average Pooling **(GAP)** under **P1**

$$p_c^{\text{gap}} = \frac{1}{H \times W} \sum_{(h,w)} \mathbf{F}_c(h,w)$$

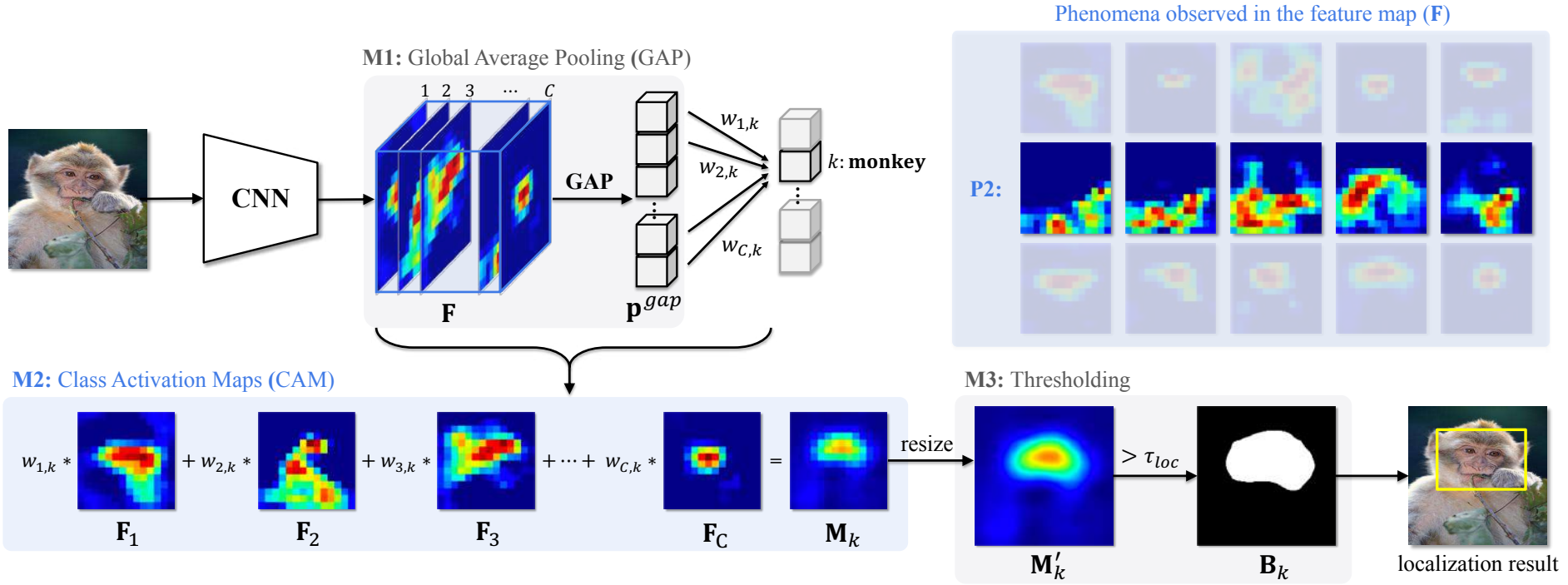
- **Solution:** Thresholded Average Pooling **(TAP)**

$$p_c^{\text{tap}} = \frac{\sum_{(h,w)} \mathbb{1}(\mathbf{F}_c(h,w) > \tau_{tap}) \mathbf{F}_c(h,w)}{\sum_{(h,w)} \mathbb{1}(\mathbf{F}_c(h,w) > \tau_{tap})}$$

Our Approach (2) Negative Weight Clamping

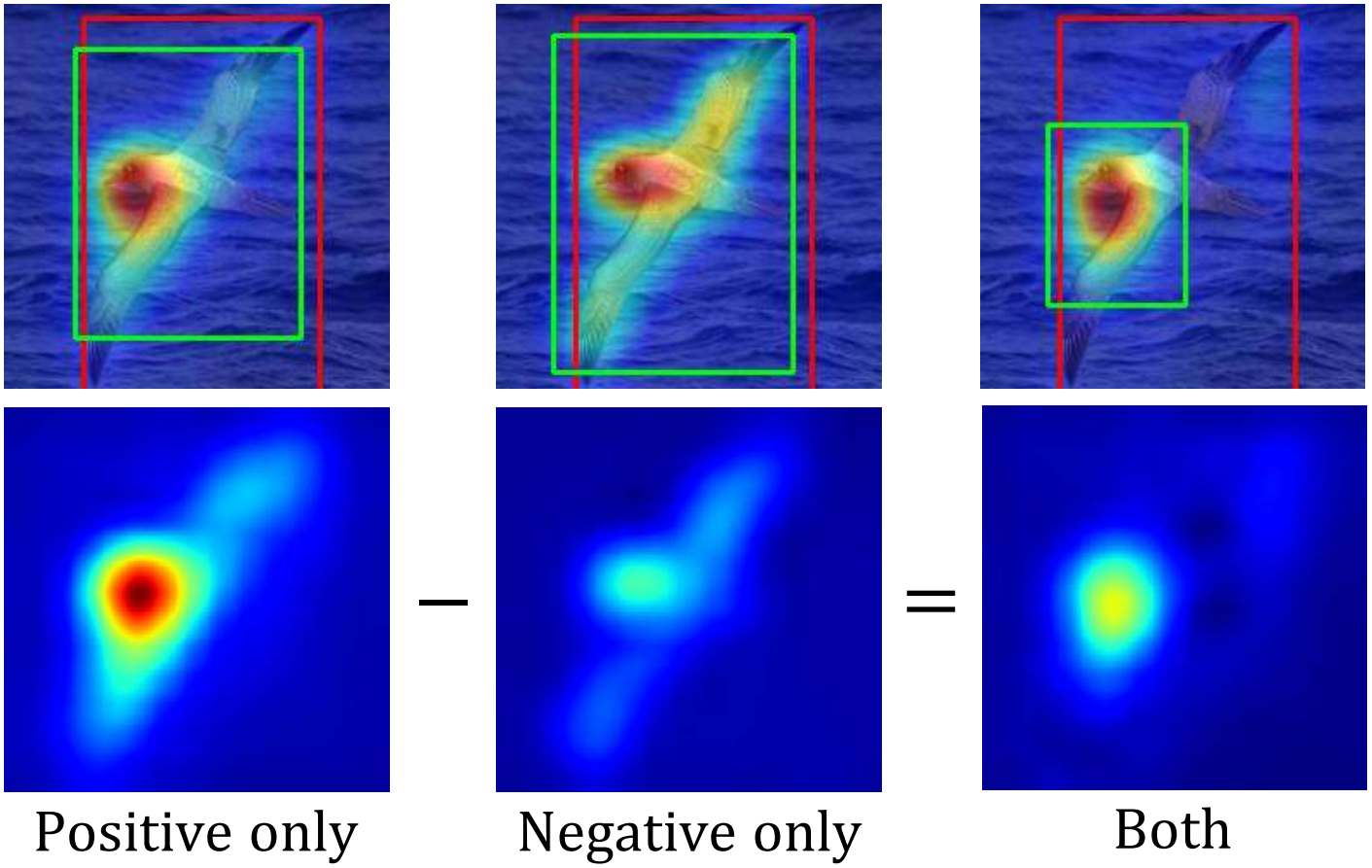
- **Problem**: Class Activation Maps **(CAM)** under **P2**

$$\mathbf{M}_k = \sum_{c=1}^{C} w_{c,k} \cdot \mathbf{F}_c$$



Our Approach (2) Negative Weight Clamping

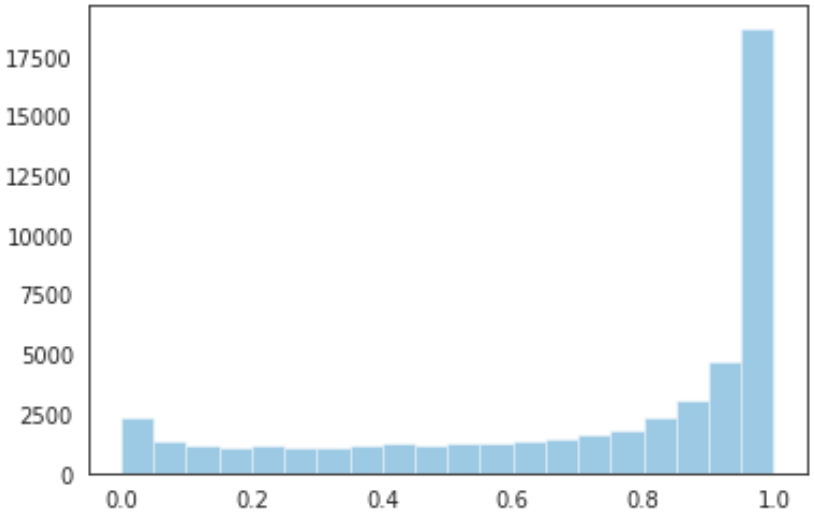
- **Problem**: Class Activation Maps **(CAM)** under **P2**



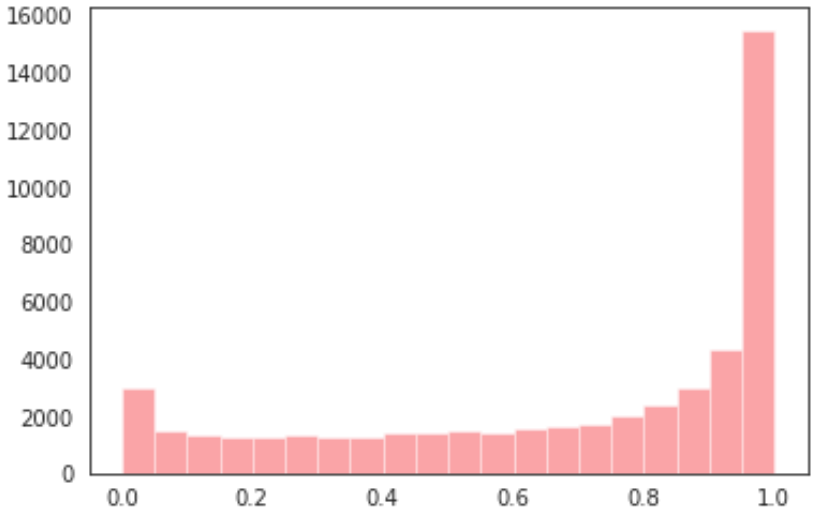
Our Approach (2) Negative Weight Clamping

- **Problem**: Class Activation Maps **(CAM)** under **P2**

IoA between the ground truth boxes and the CAMs



Positive weights



Negative weights

Our Approach (2) Negative Weight Clamping

- **Problem**: Class Activation Maps **(CAM)** under **P2**

$$\mathbf{M}_k = \sum_{c=1}^{C} w_{c,k} \cdot \mathbf{F}_c$$

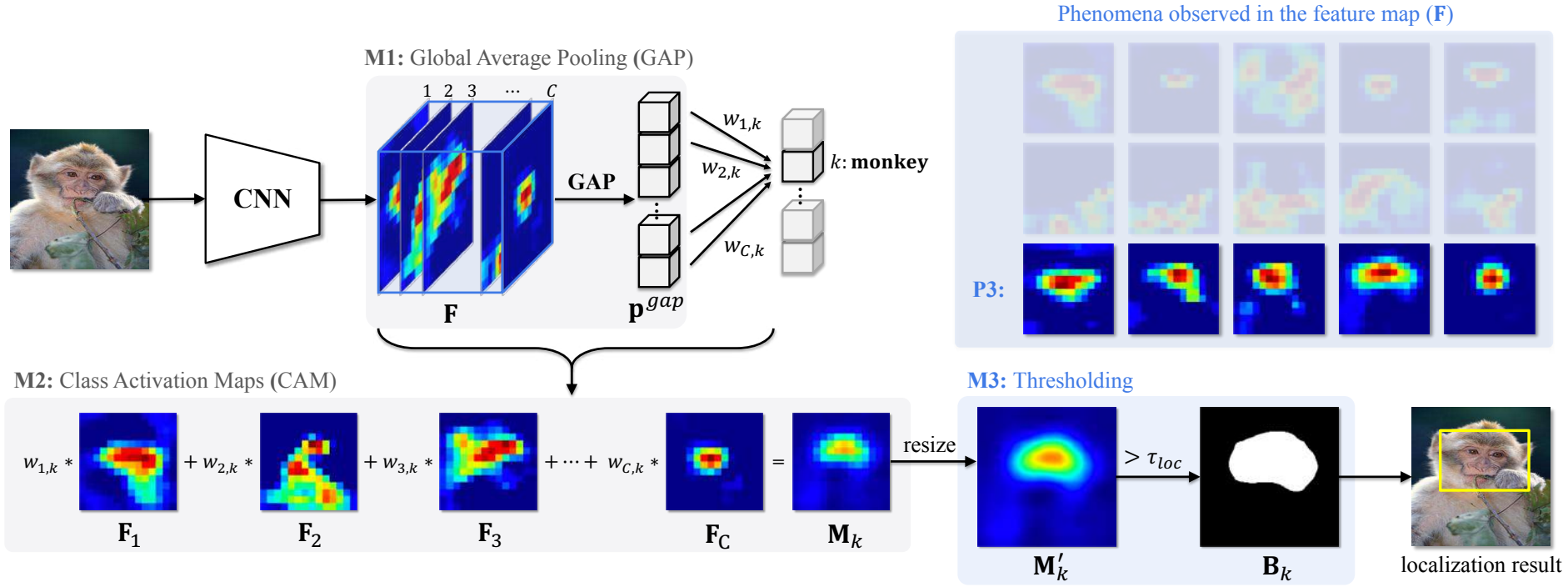
- **Solution**: Negative Weight Clamping **(NWC)**

$$\mathbf{M}_k = \sum_{c=1}^{C} \mathbb{1}(w_{c,k} > 0) \cdot w_{c,k} \cdot \mathbf{F}_c$$

Our Approach (3) Percentile as a Thresholding Standard

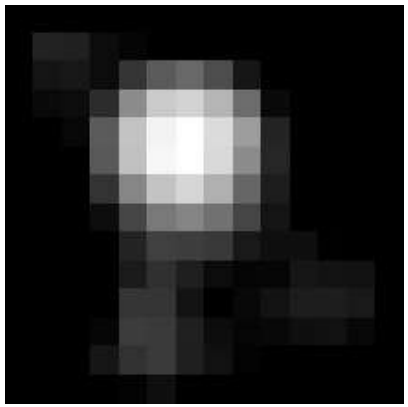
- **Problem:** Maximum as a Standard **(MaS)** under **P3**

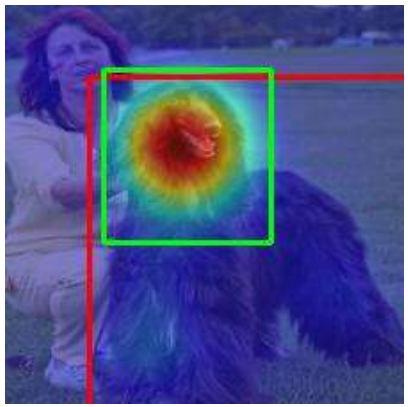
$$\tau_{loc} = \theta_{loc} \cdot \max \mathbf{M}'_k$$

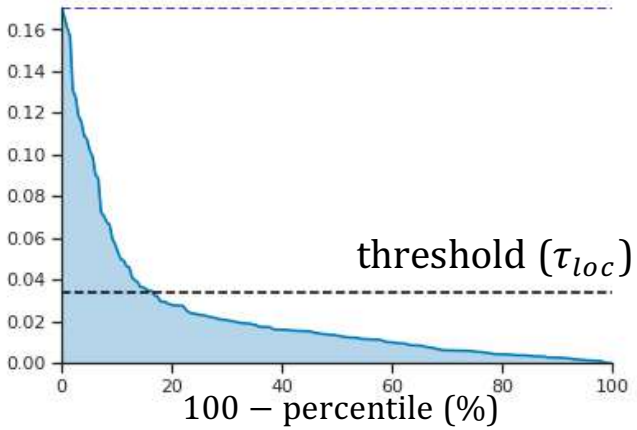


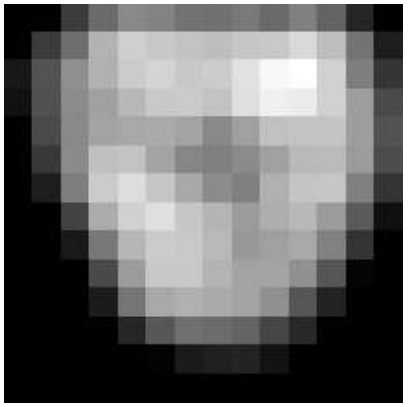
Our Approach (3) Percentile as a Thresholding Standard

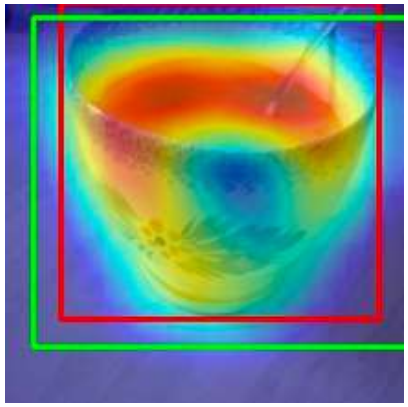
- **Problem:** Maximum as a Standard **(MaS)** under **P3**

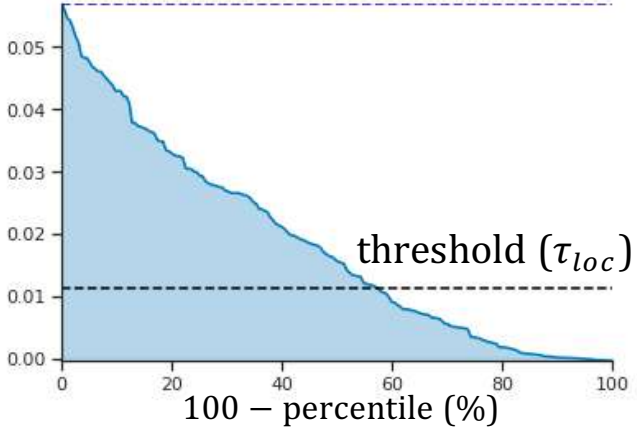












Num of channels (activation $> \tau_{0.8}$)

Result with CAM

CAM values (descending order)

Our Approach (3) Percentile as a Thresholding Standard

- **Problem:** Maximum as a Standard **(MaS)** under **P3**

$$\tau_{loc} = \theta_{loc} \cdot \max \mathbf{M}'_k$$

- **Solution:** Percentile as a Standard **(PaS)**

$$\tau_{loc} = \theta_{loc} \cdot \mathrm{per}_i(\mathbf{M}'_k)$$

Experimental Setting

- Backbone: ResNet50-SE
- Batch size: 210
- Input size: 384×384
- Random crop size: 336×336
- TAP threshold (τ_{tap}): 0.05
- PaS percentile (i): 98

Results on Validation Set

- Results with different components

Method	CRF	PaS	NWC	TAP	Peak IoU
Baseline					0.5254
	✓				0.5461
+ Ours	✓	✓			0.5563
	✓	✓	✓		0.5881
	✓	✓	✓	✓	**0.6370**

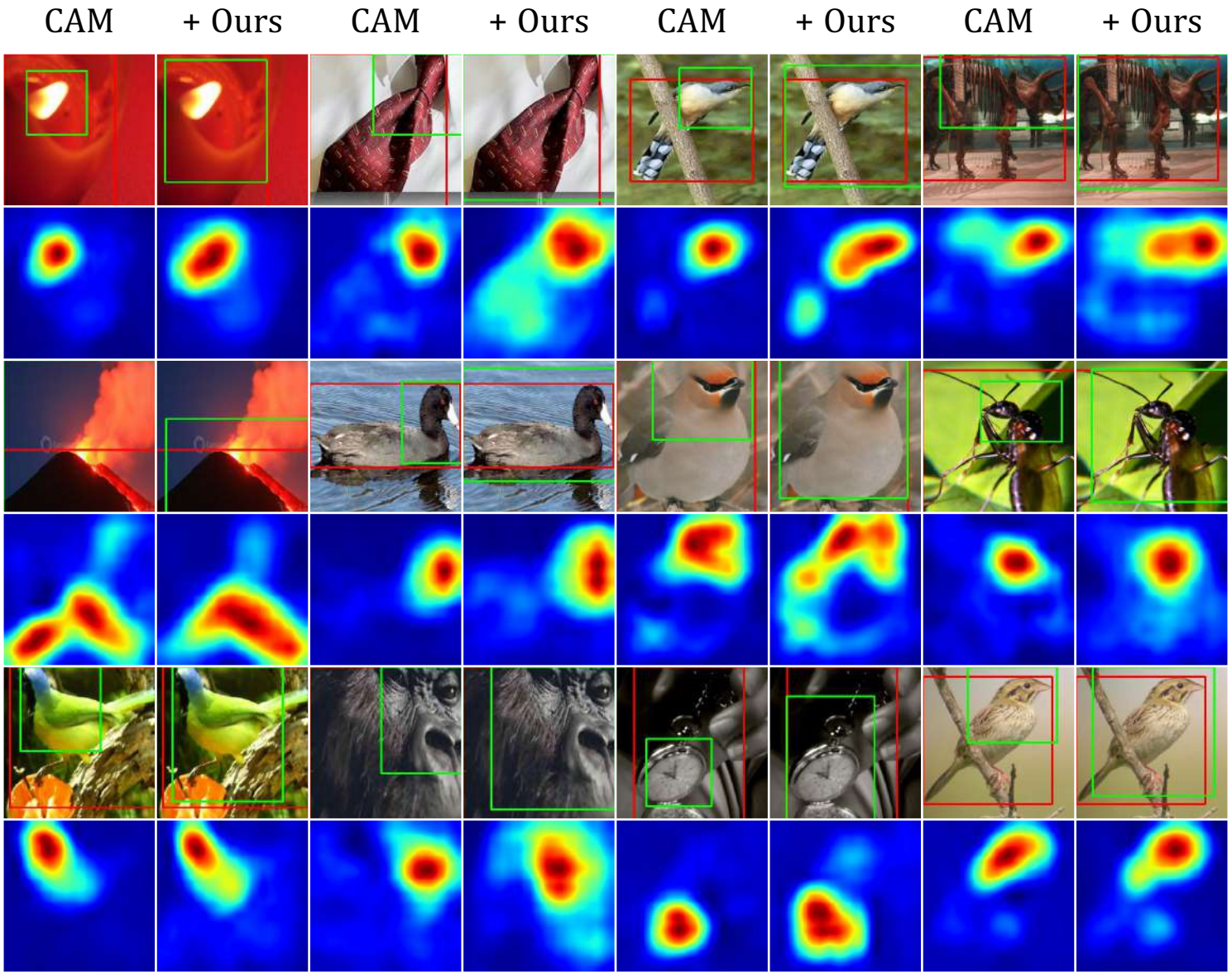
- To preserve the details of masks, we also applied a fully connected CRF.
- The performance gradually improves as each component is added.

Leaderboard

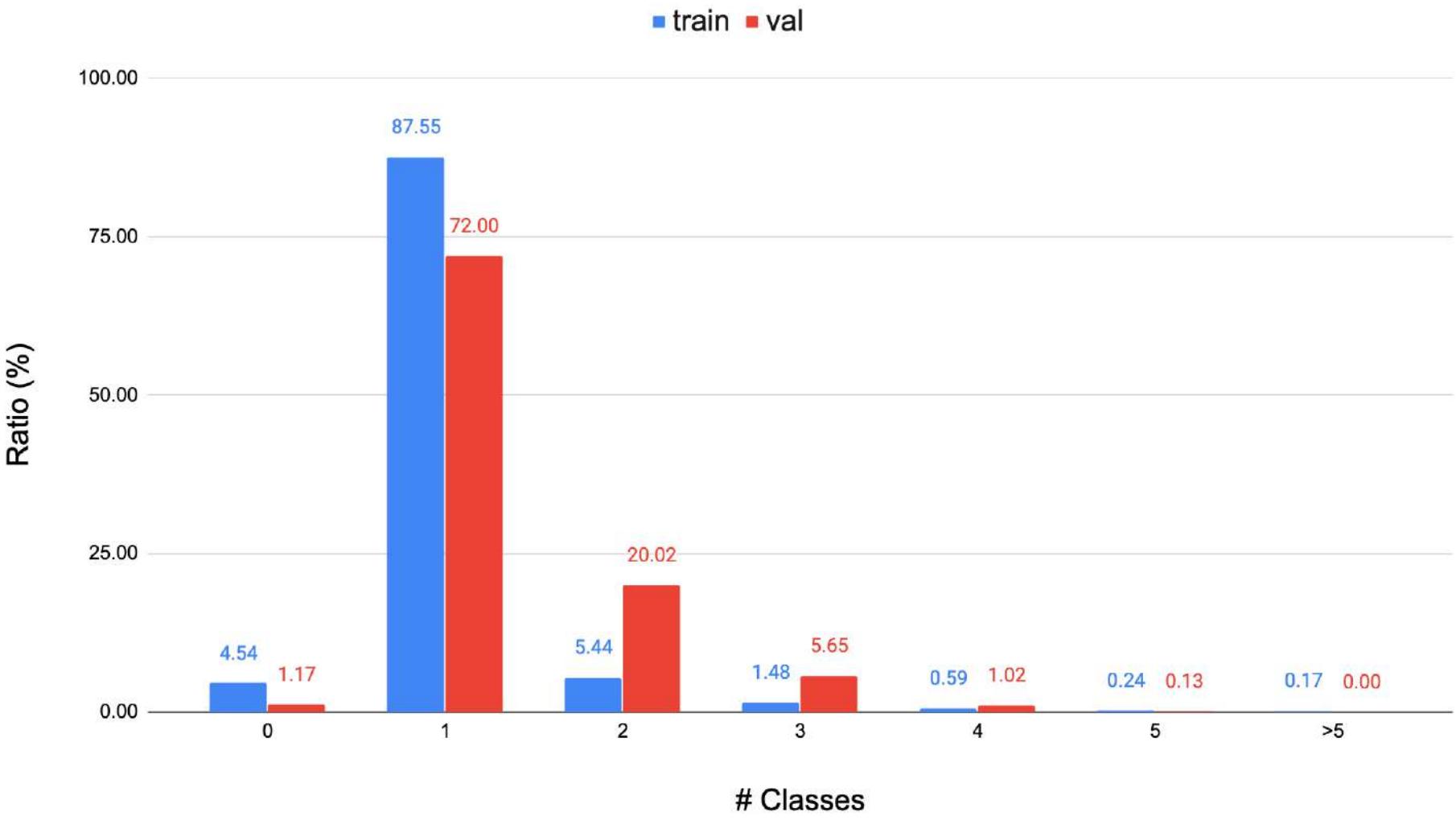
- Track 3: Weakly Supervised Object Localization

Rank	Participant Team	Peak IoU
1	**SNUVL (Ours)**	**0.63**
2	BJTU-Mepro-MIC	0.62
3	LEAP Group@PCA Lab	0.61
4	chohk (wsol_aug)	0.53
5	TEN	0.48

Qualitative Results

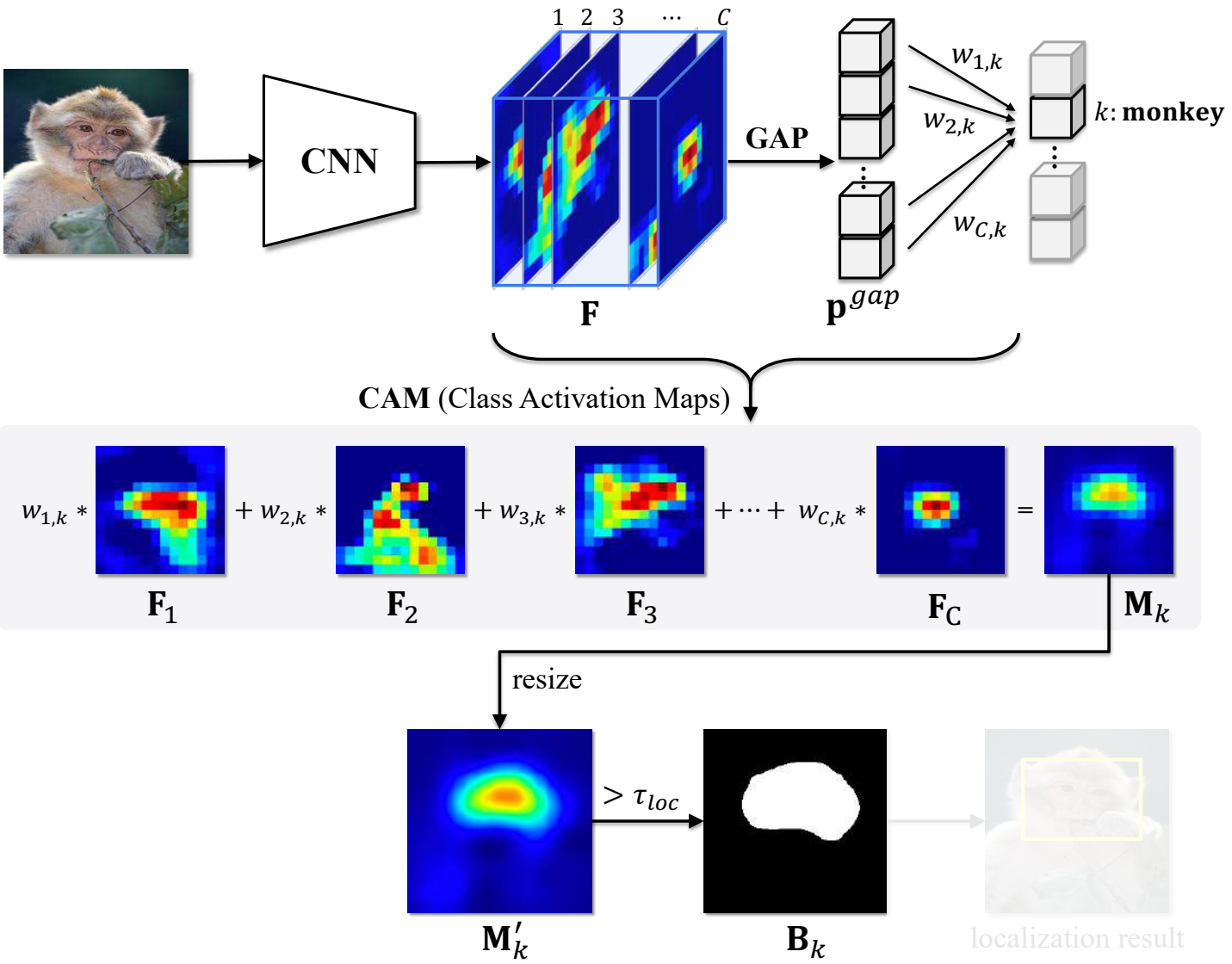


Expansion to Track 1



Expansion to Track 1

Class Activation Mapping (CAM) for Track 1



Leaderboard

- Track 1: Weakly Supervised Semantic Segmentation

Rank	Participant Team	Mean IoU
1	cvl	45.18
2	**SNUVL (Ours)**	**37.73**
3	UCU & SoftServe	37.34
4	IOnlyHaveSevenDays	36.24
5	play-njupt	31.90

Thank You!