On Train-Test Class Overlap and Detection for Image Retrieval

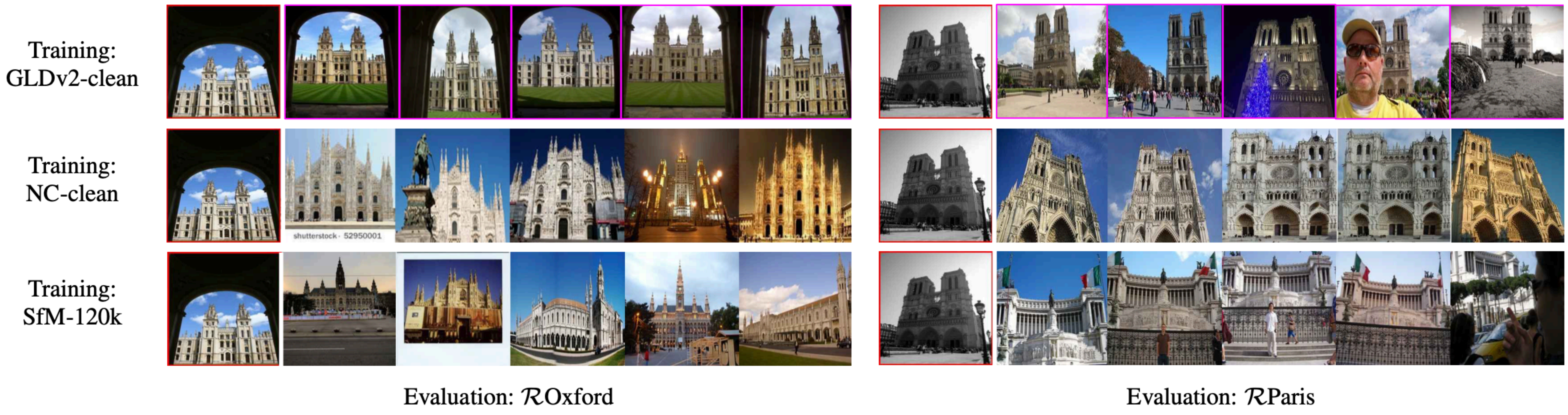
Chull Hwan Song[1] Jooyoung Yoon[1] Taebaek Hwang[1] Shunghyun Choi[1] YeongHyeon Gu[2*] Yannis Avrithis[3]

[1]Dealicious Inc. [2]Sejong University [3]Institute of Advanced Research on Artificial Intelligence (IARAI)

Introduction

- **Background:** Importance of non-overlapping training and evaluation sets in image retrieval.
- **Problem Statement:** Existing methods do not adequately address class overlap issues in datasets.
- **Objective:** Introduce RGLDv2-clean and CiDeR for effective image retrieval.

Main Contributions



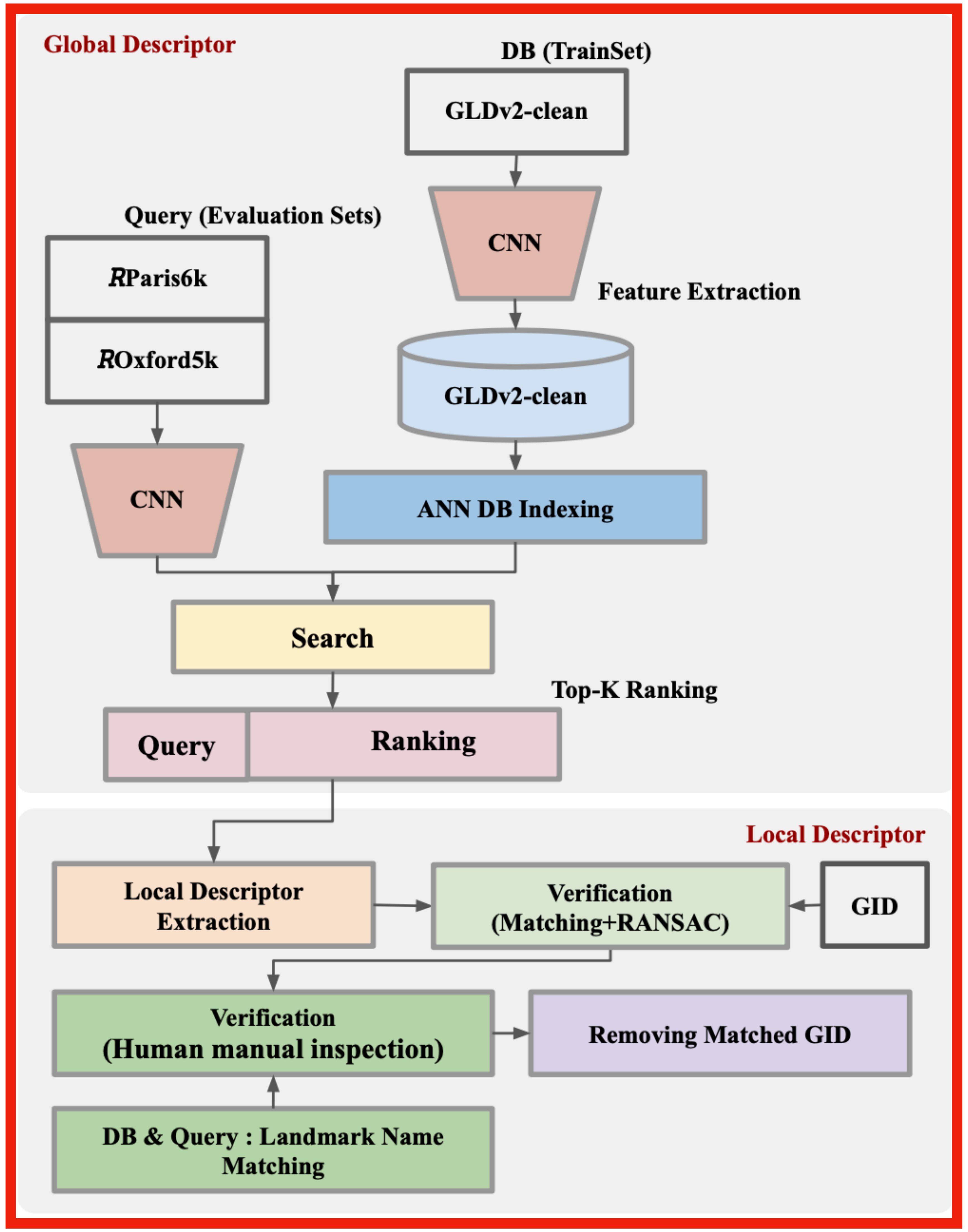
- Key Points: Overlap between GLDv2-clean and evaluation sets(ROxford and RPar)

Data Cleaning Process

- Steps: Identify, remove overlaps, verify
- Table: Statistics before and after cleaning

EVAL	#EVAL IMG	#DUPL EVAL	#DUPL GLDV2 GID	#DUPL GLDV2 IMG
$\mathcal{R}$Par	70	36 (51%)	11	1,227
$\mathcal{R}$Oxf	70	38 (54%)	6	315
TEXT			1	23
TOTAL	140	74	18	1,565

TRAINING SET	#IMAGES	#CATEGORIES
NC-clean	27,965	581
SfM-120k	117,369	713
GLDv2-clean	1,580,470	81,313
$\mathcal{R}$GLDv2-clean (ours)	1,578,905	81,295

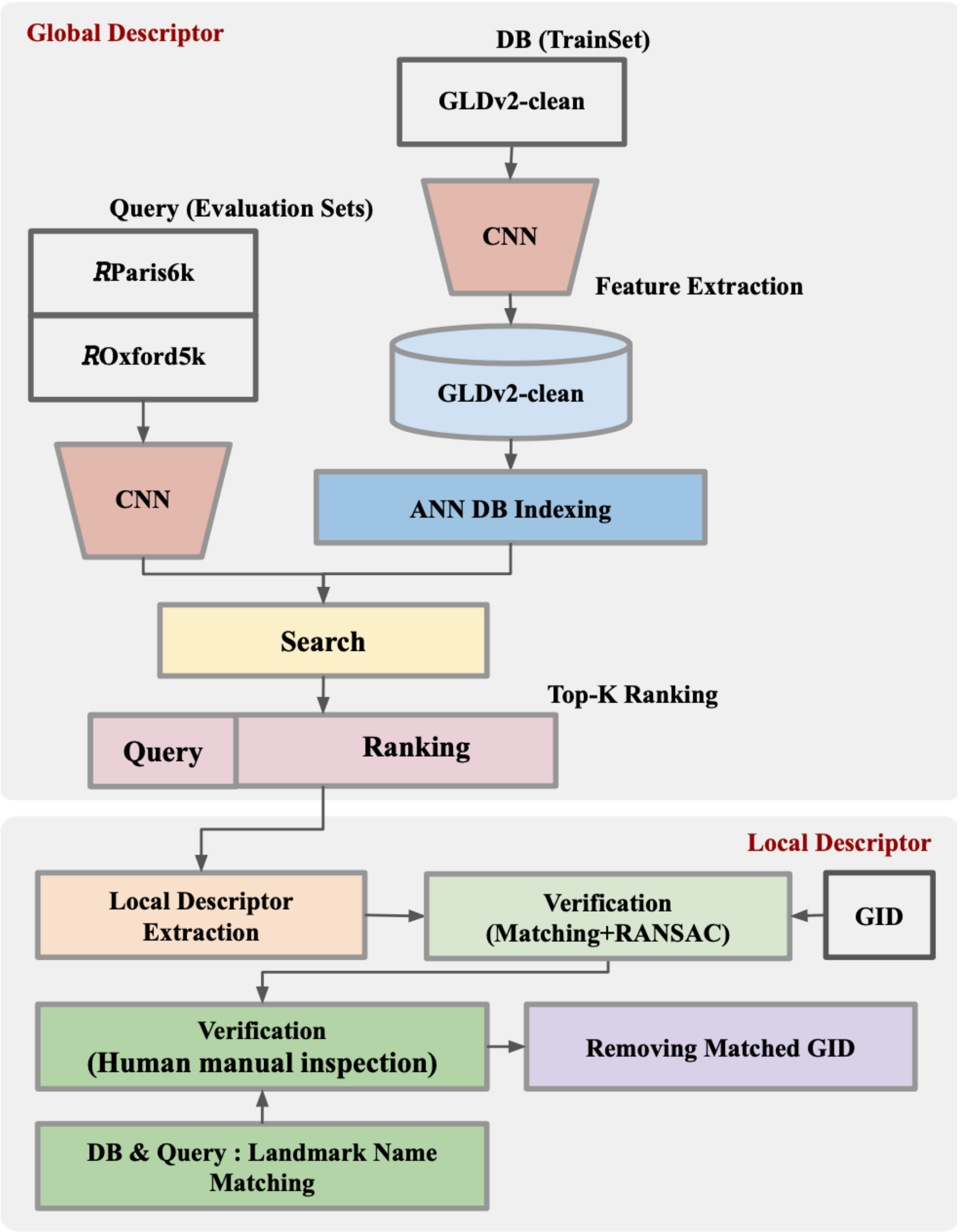


Data Cleaning Process

- Steps: Identify, remove overlaps, verify
- Table: Statistics before and after cleaning

EVAL	#EVAL IMG	#DUPL EVAL	#DUPL GLDV2 GID	#DUPL GLDV2 IMG
$\mathcal{R}$Par	70	36 (51%)	11	1,227
$\mathcal{R}$Oxf	70	38 (54%)	6	315
TEXT			1	23
TOTAL	140	74	18	1,565

TRAINING SET	#IMAGES	#CATEGORIES
NC-clean	27,965	581
SfM-120k	117,369	713
GLDv2-clean	1,580,470	81,313
$\mathcal{R}$GLDv2-clean (ours)	1,578,905	81,295

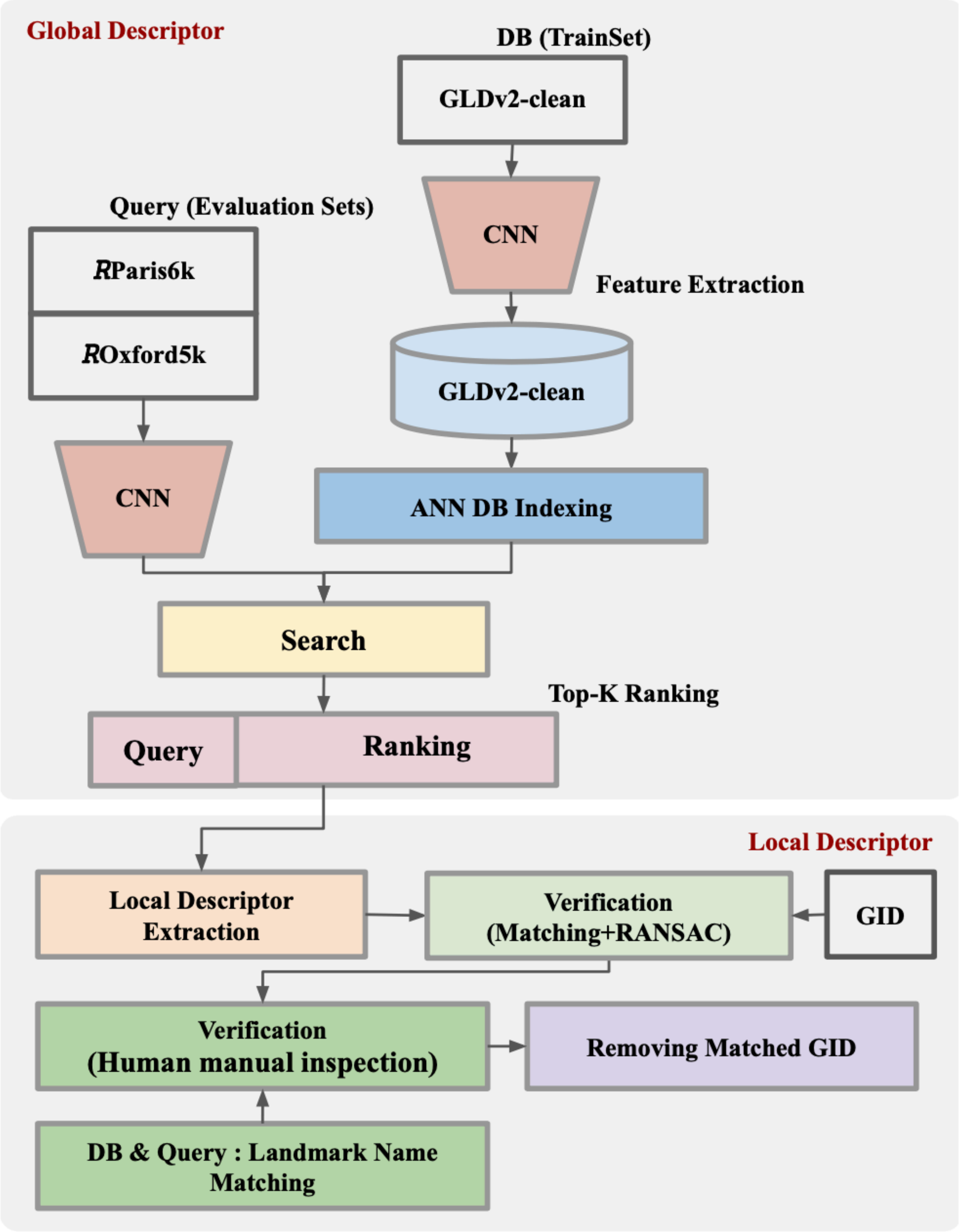


Data Cleaning Process

- Steps: Identify, remove overlaps, verify
- Table: Statistics before and after cleaning

EVAL	#EVAL IMG	#DUPL EVAL	#DUPL GLDV2 GID	#DUPL GLDV2 IMG
$\mathcal{R}$Par	70	36 (51%)	11	1,227
$\mathcal{R}$Oxf	70	38 (54%)	6	315
TEXT			1	23
TOTAL	140	74	18	1,565

TRAINING SET	#IMAGES	#CATEGORIES
NC-clean	27,965	581
SfM-120k	117,369	713
GLDv2-clean	1,580,470	81,313
$\mathcal{R}$GLDv2-clean (ours)	1,578,905	81,295

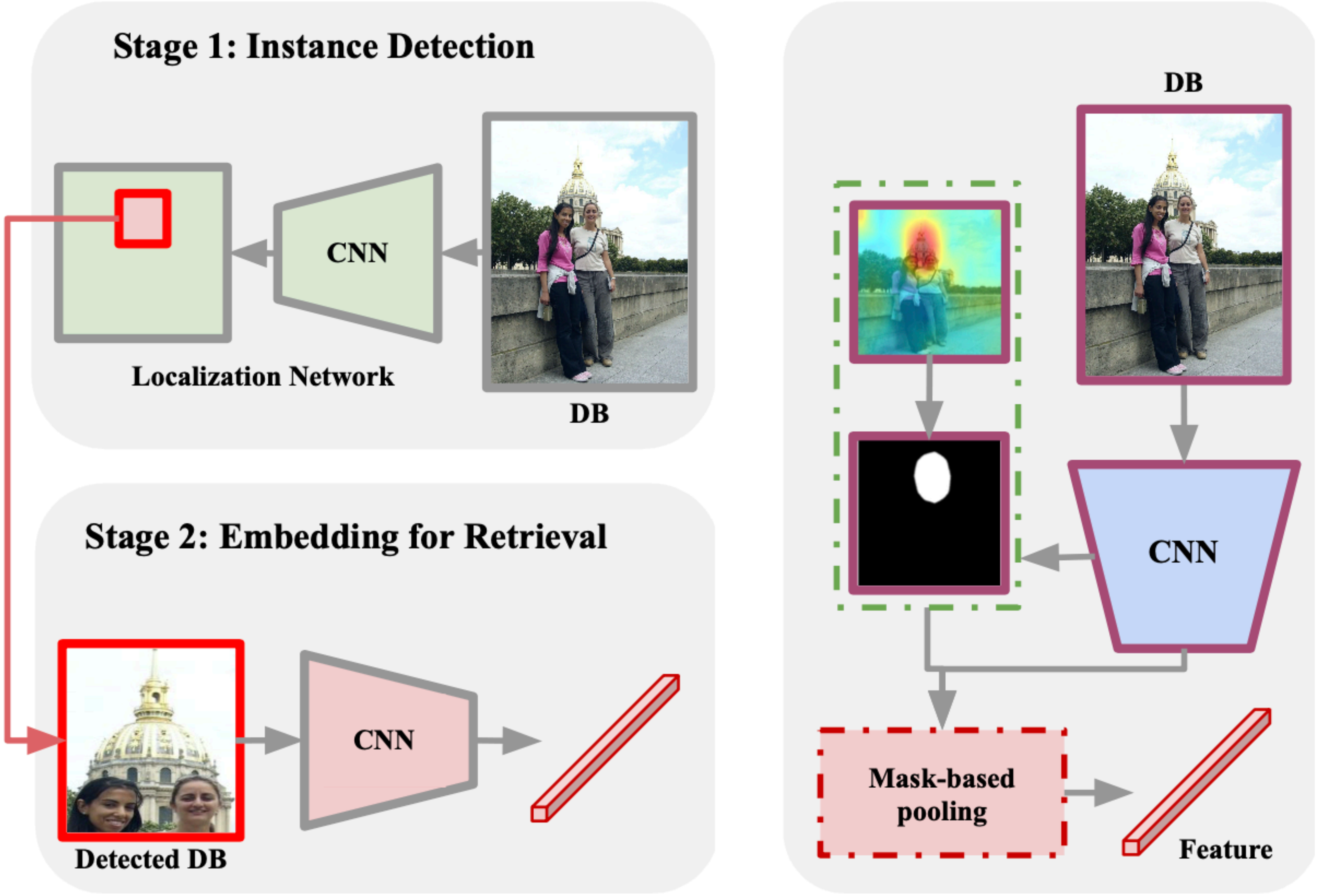


Performance Impact

- Performance comparison on GLDv2-clean vs RGLDv2-clean
- Significant drop in performance on cleaned dataset

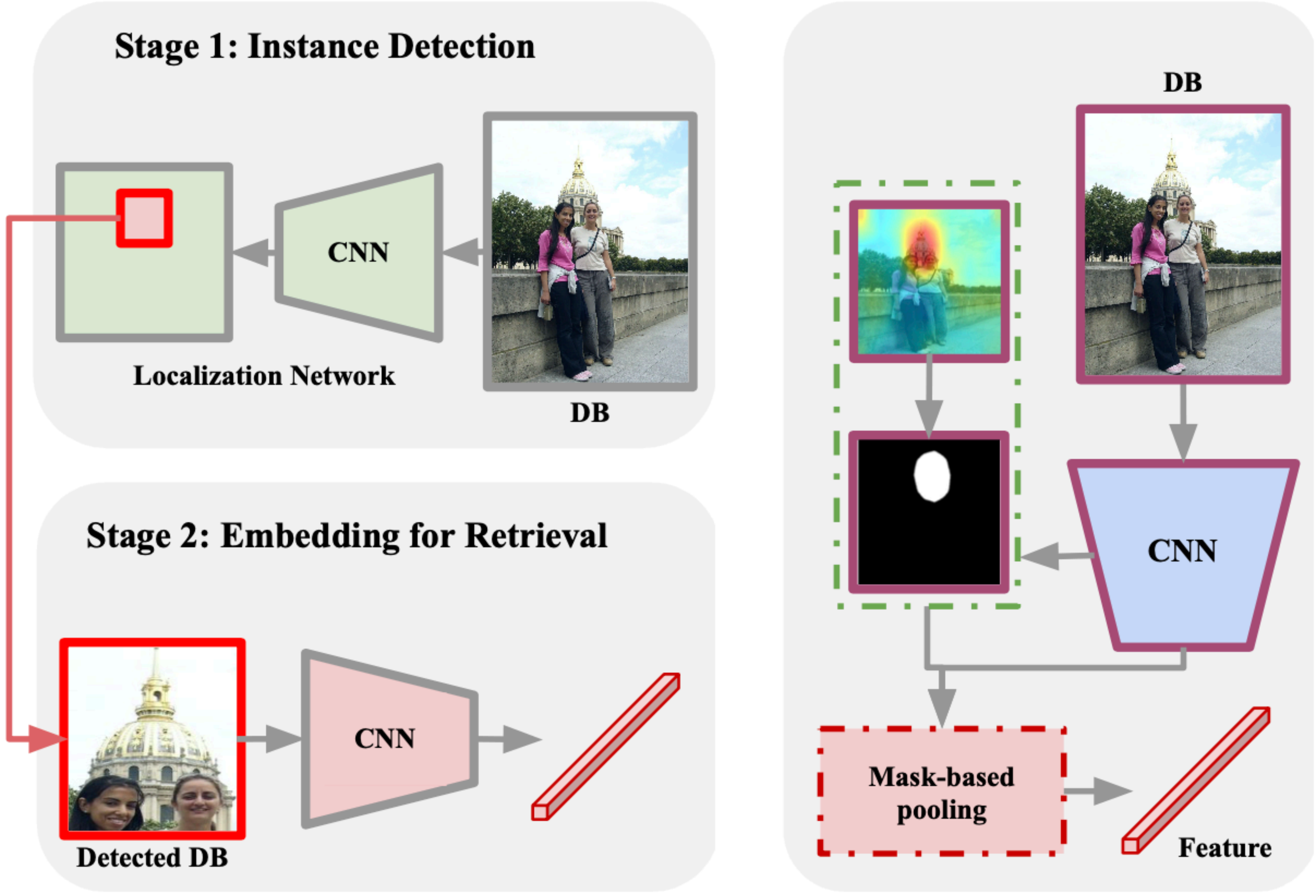
METHOD	TRAIN SET	BASE		MEDIUM				HARD				MEAN	DIFF
		Ox5k	Par6k	$\mathcal{R}$Oxf		$\mathcal{R}$Par		$\mathcal{R}$Oxf		$\mathcal{R}$Par			
		mAP	mAP	mAP	mP@10	mAP	mP@10	mAP	mP@10	mAP	mP@10		
Yokoo *et al*. [46]	GLDv2-clean	91.9	94.5	72.8	86.7	84.2	95.9	49.9	62.1	69.7	88.4	79.5	-5.4
Yokoo *et al*. [60]†	$\mathcal{R}$GLDv2-clean	86.1	93.9	64.5	81.0	84.1	95.4	35.6	51.5	68.7	86.4	74.1	
SOLAR [58]	GLDv2-clean	–	–	79.7	–	88.6	–	60.0	–	75.3	–	75.9	-8
SOLAR [27]†	$\mathcal{R}$GLDv2-clean	90.6	94.4	70.8	84.6	84.1	95.4	48.0	62.3	68.7	86.4	67.9	
GLAM [46]	GLDv2-clean	94.2	95.6	78.6	88.2	88.5	97.0	60.2	72.9	76.8	93.4	83.4	-4.1
GLAM [46]‡	$\mathcal{R}$GLDv2-clean	90.9	94.1	72.2	84.7	83.0	95.0	49.6	61.6	65.6	87.6	79.3	
DOLG [47]	GLDv2-clean	–	–	78.8	–	87.8	–	58.0	–	74.1	–	74.7	-7.4
DOLG [59]†	$\mathcal{R}$GLDv2-clean	88.3	93.9	70.8	85.3	83.2	95.4	47.4	60.0	67.9	87.4	67.3	
Token [58]	GLDv2-clean	–	–	82.3	–	75.6	–	66.6	–	78.6	–	75.8	-18.2
Token [58]†	$\mathcal{R}$GLDv2-clean	84.3	90.0	61.4	76.4	75.8	94.0	36.9	55.2	54.4	81.0	57.6	

Main Contributions 2



(a) Two-stage (b) One-stage

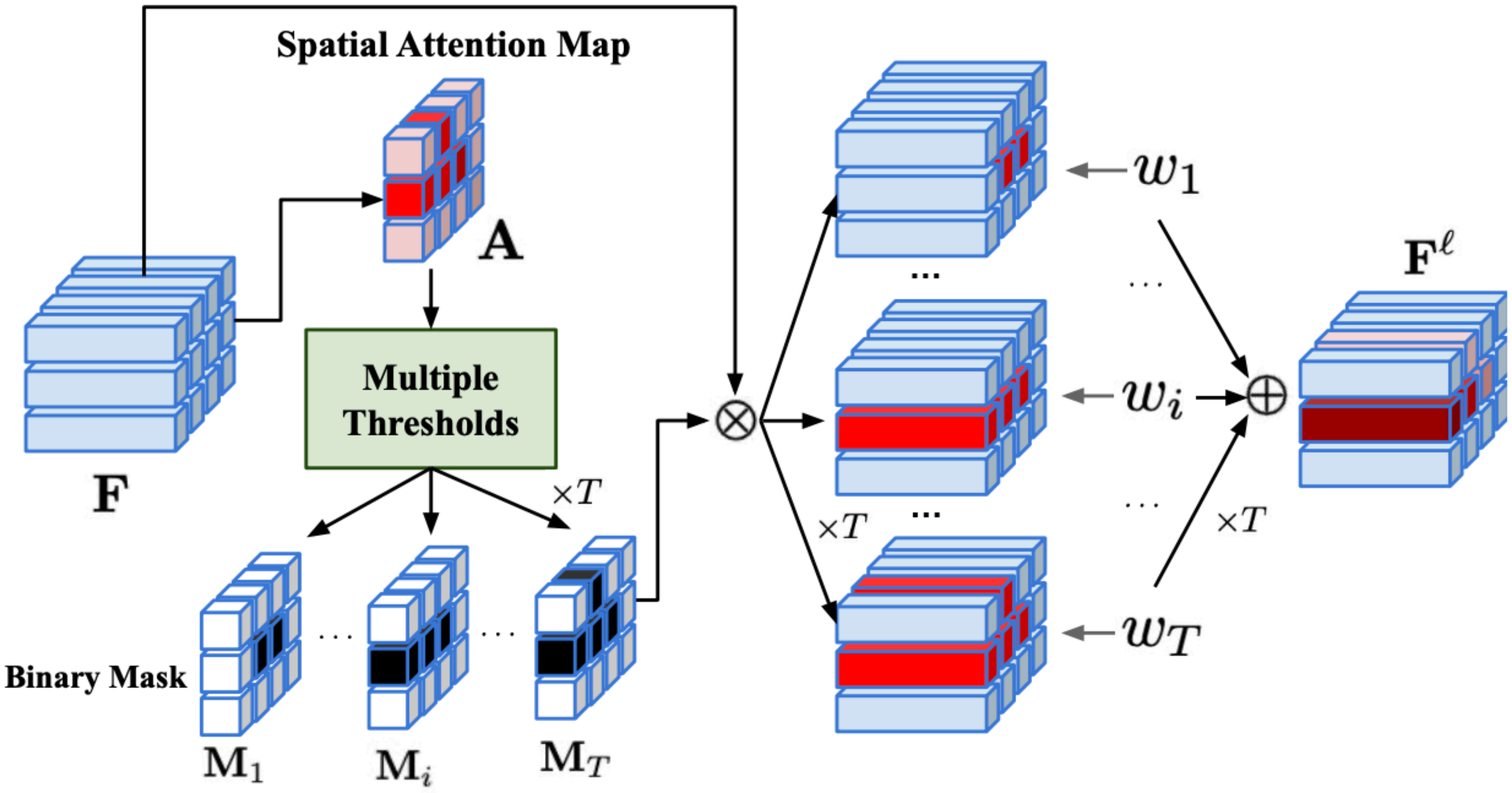
Main Contributions 2



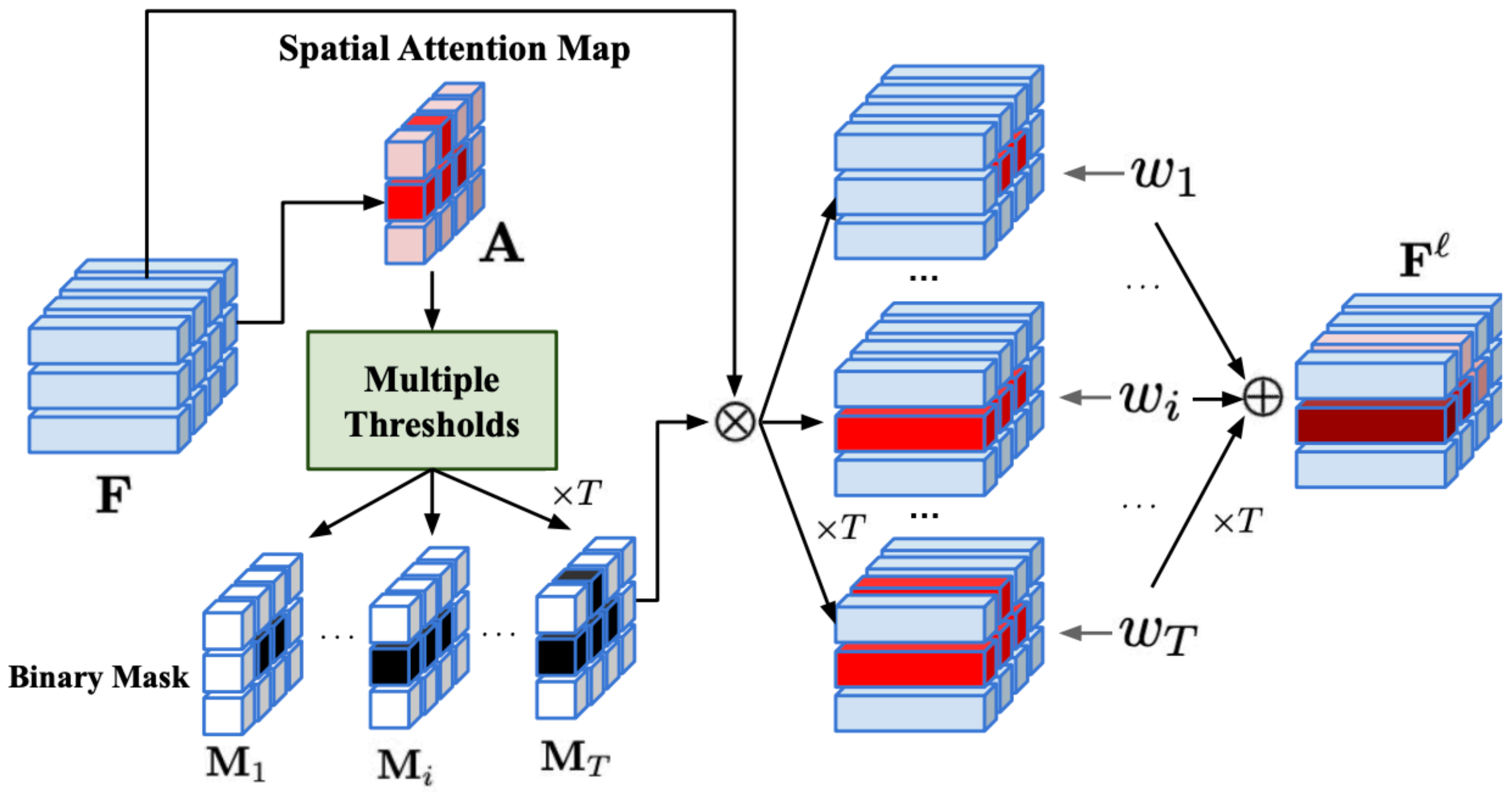
(a) Two-stage (b) One-stage

Attentional Localization

- How spatial attention maps and masks are used

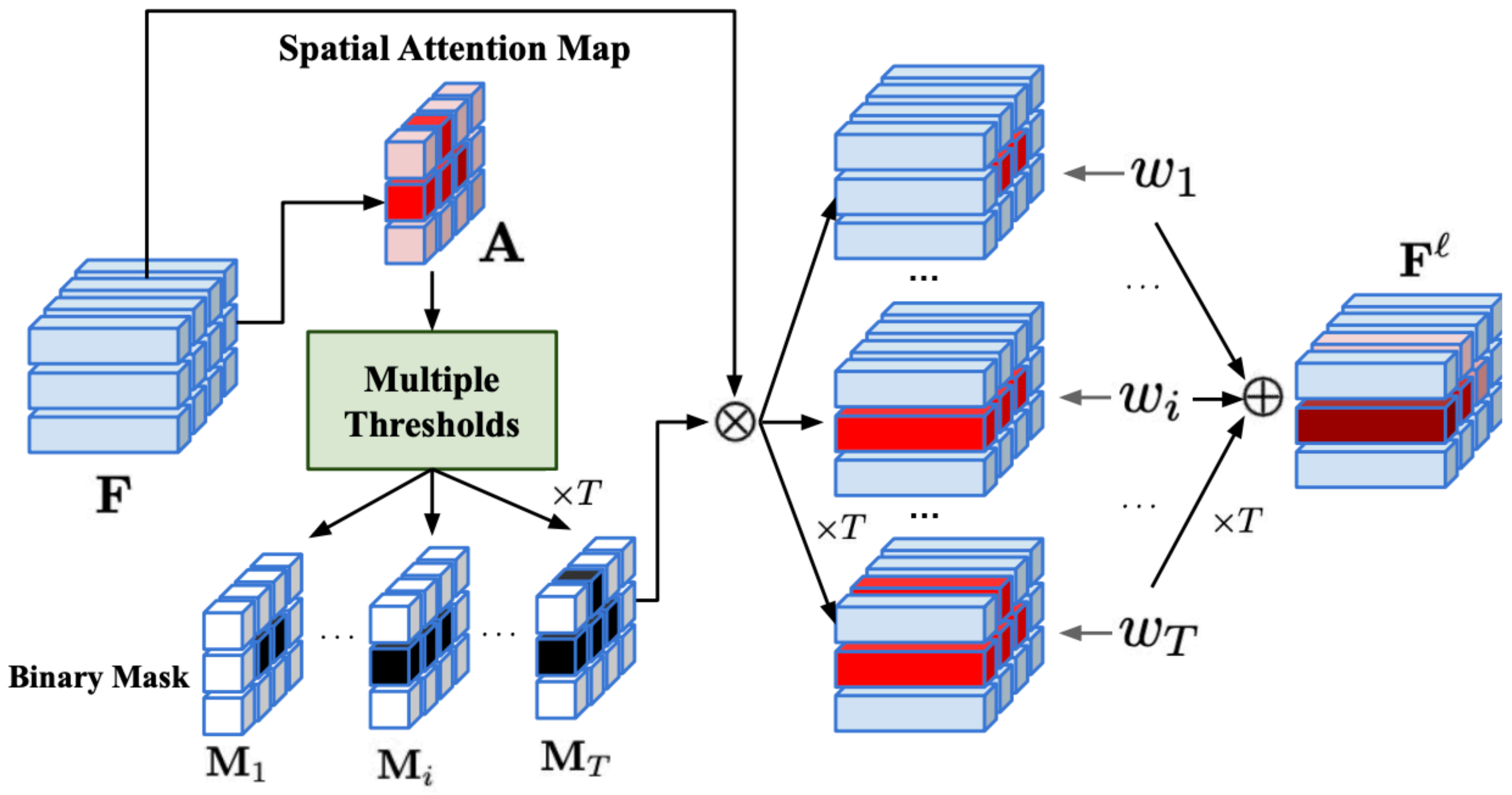


Attentional Localization



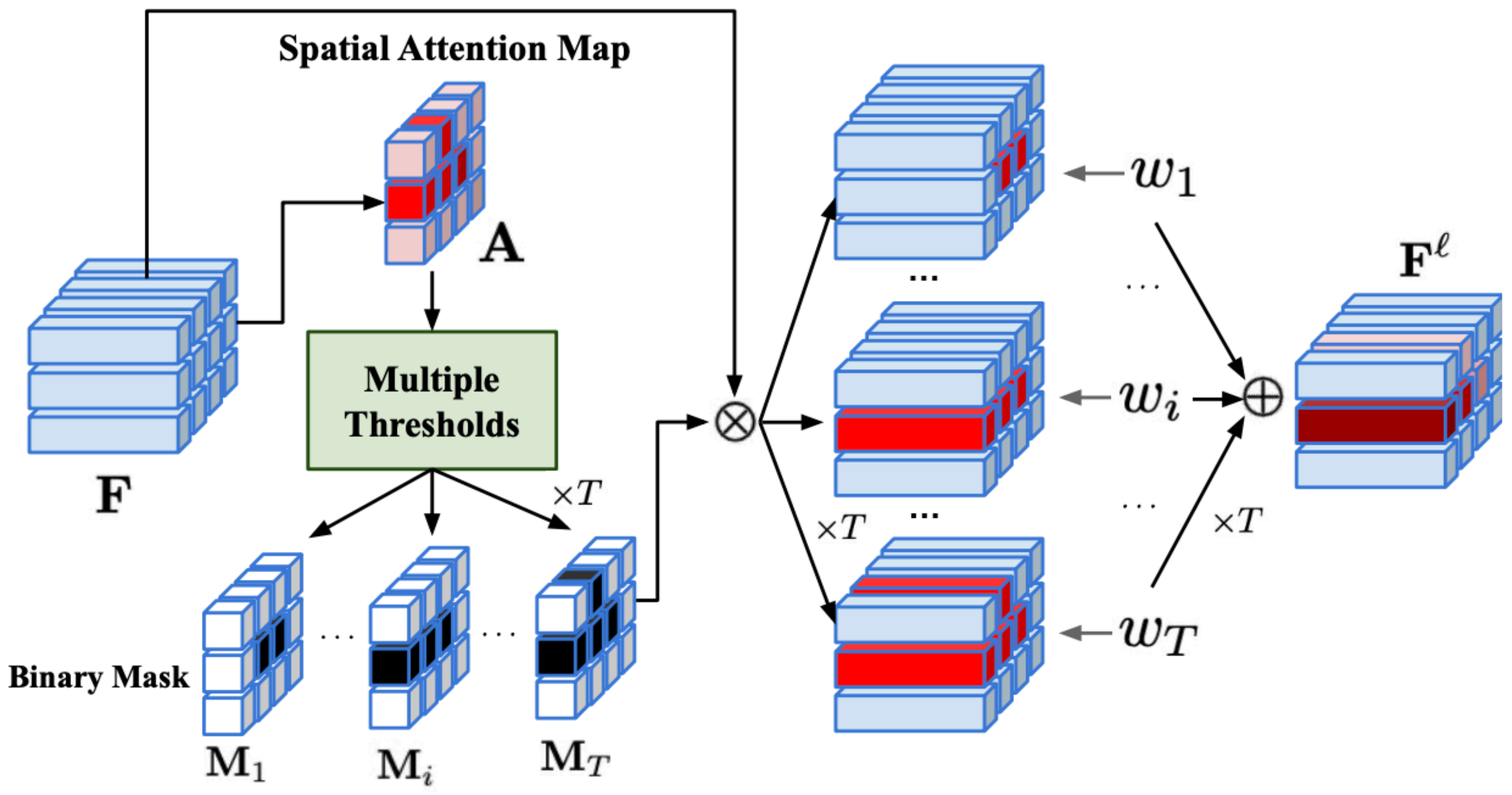
$$A = \eta(\zeta(f^\ell(\mathbf{F}))) \in \mathbb{R}^{w \times h}$$

Attentional Localization



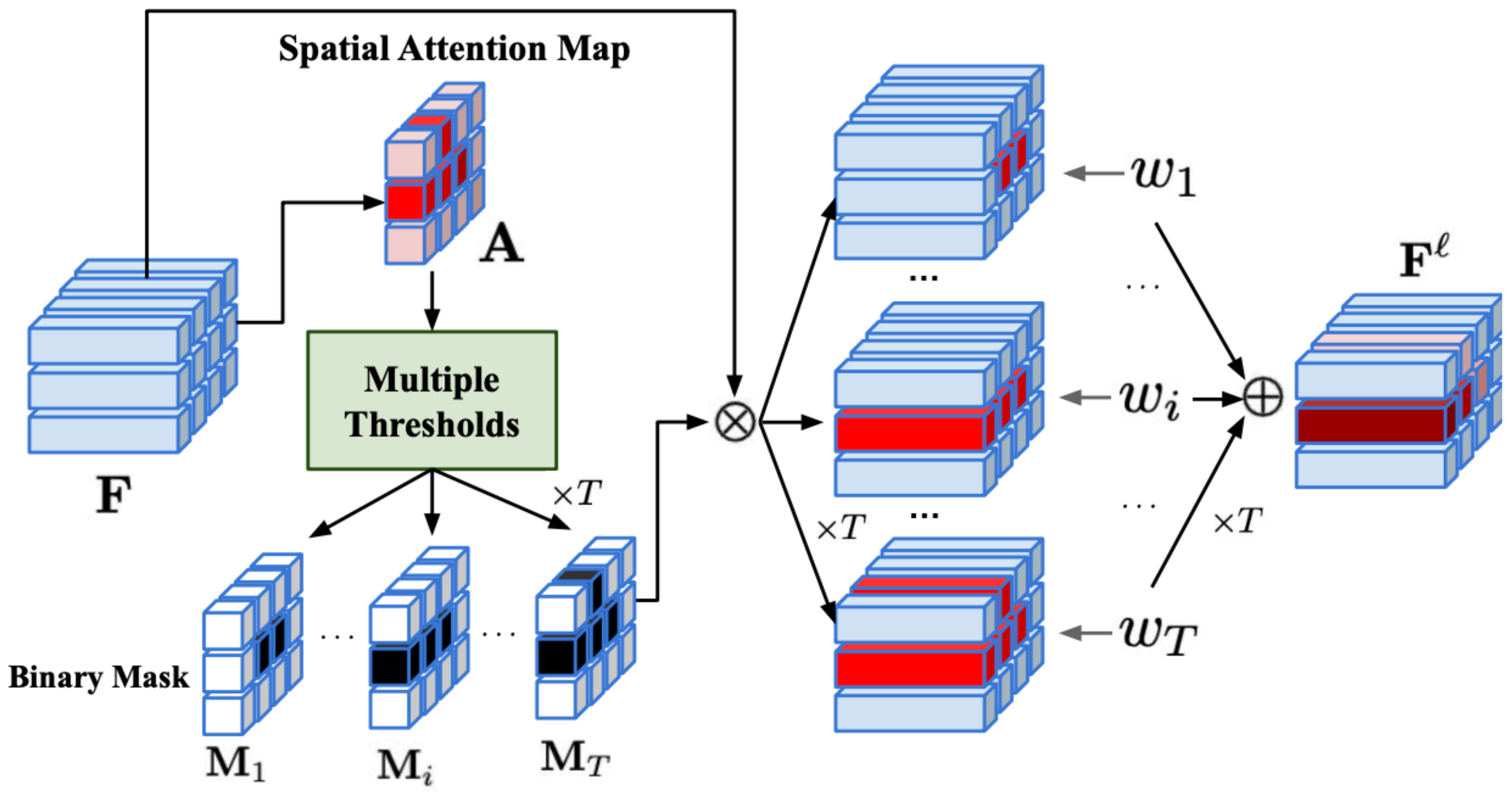
$$\eta(X) := \frac{X - \min X}{\max X - \min X} \in \mathbb{R}^{w \times h}$$

Attentional Localization



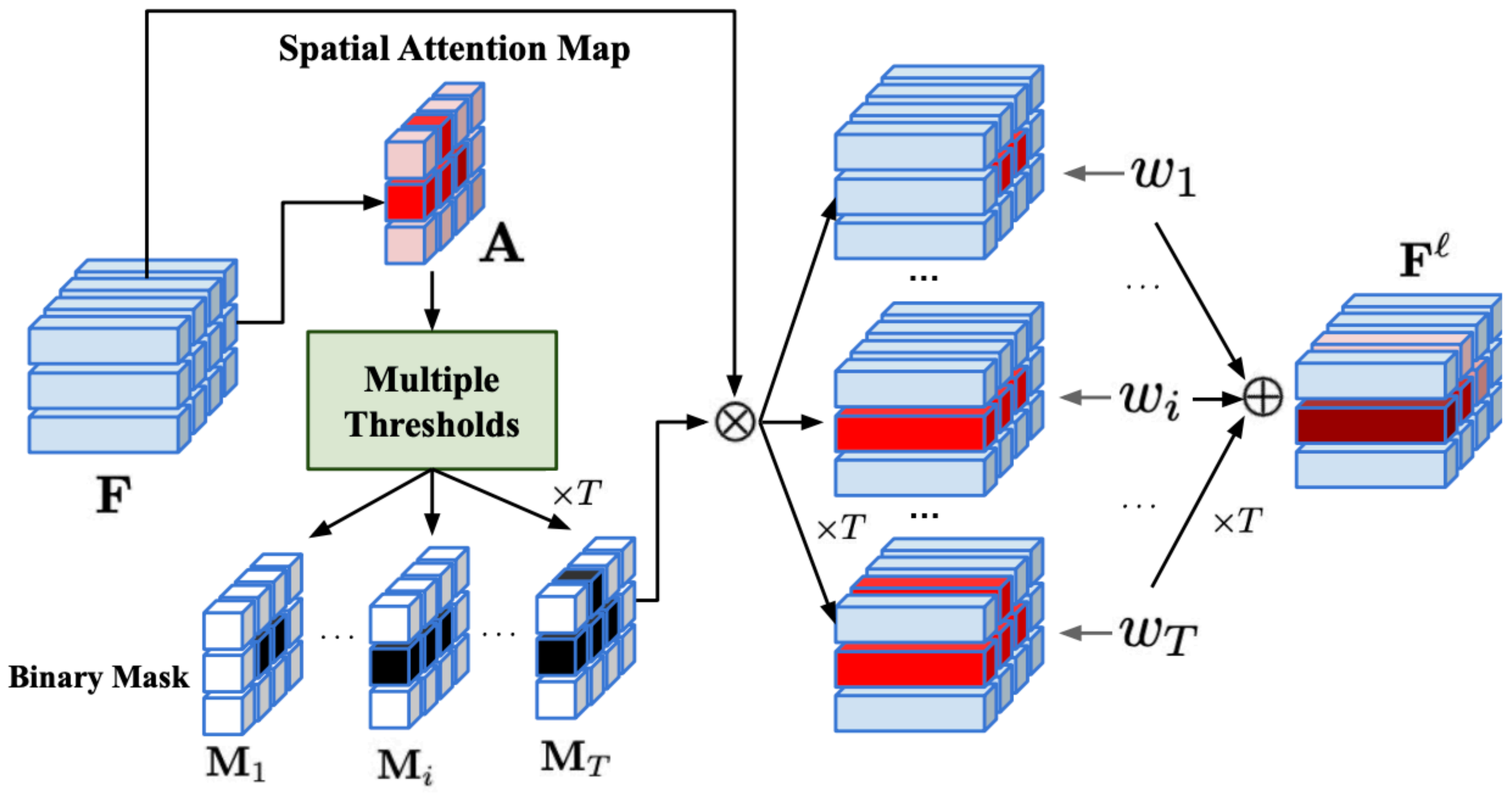
$$M_i(\mathbf{p}) = \begin{cases} \beta, & \text{if } A(\mathbf{p}) < \tau_i \\ 1, & \text{otherwise} \end{cases}$$

Attentional Localization



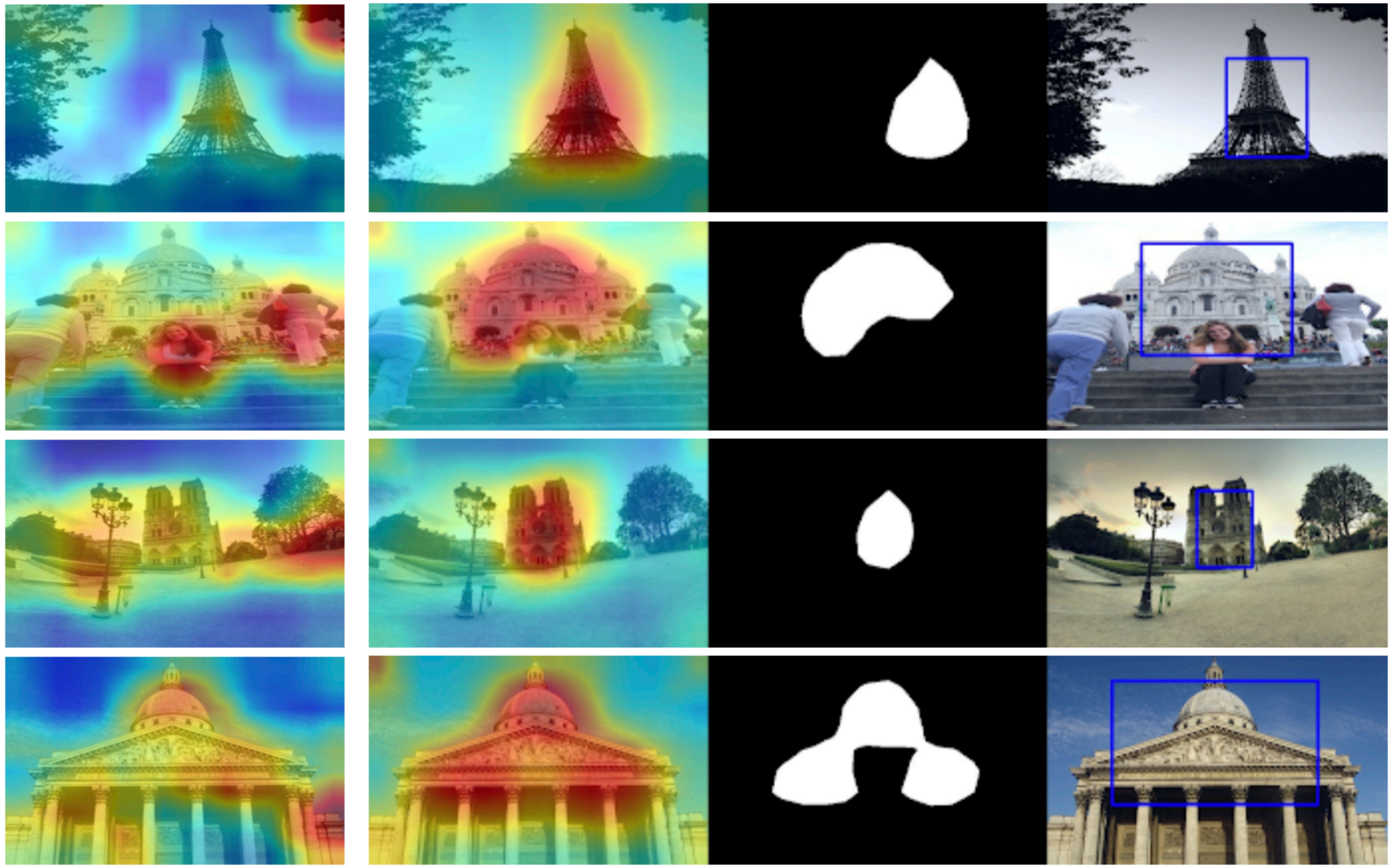
$$\mathbf{F}^{\ell} = \mathcal{H}(M_1 \odot \mathbf{F}, \ldots, M_T \odot \mathbf{F}) \in \mathbb{R}^{w \times h \times d}$$

Attentional Localization



$$\mathcal{H}(\mathbf{F}_1, \ldots, \mathbf{F}_T) = \frac{w_1\mathbf{F}_1 + \cdots + w_T\mathbf{F}_T}{w_1 + \cdots + w_T}$$

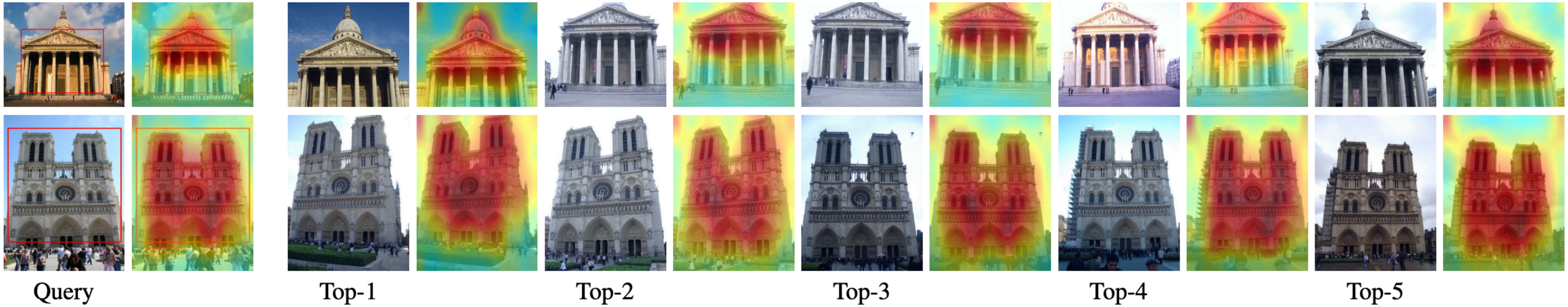
Visual Comparison



(a) A, pre-trained (b) A, ours (c) Mask M_i (d) Bounding box

Visualization

- Focus on relevant objects, ignoring background.



Performance Comparison

- Superior performance on various datasets

Method	Train Set	Net	Pooling	Loss	FT	E2E	Self	Dim	Base		$\mathcal{R}$Medium		$\mathcal{R}$Hard		Mean
									Oxf5k	Par6k	$\mathcal{R}$Oxf	$\mathcal{R}$Par	$\mathcal{R}$Oxf	$\mathcal{R}$Par	
Local Descriptors															
HesAff-rSIFT-ASMK⋆+SP [34]	SfM-120k	R50	–	–	✓	–	–	–	–	–	60.6	61.4	36.7	35.0	–
DELF-ASMK⋆+SP [34]	SfM-120k	R50	–	CLS	✓	–	–	–	–	–	**67.8**	**76.9**	**43.1**	**55.4**	–
Local Descriptors+D2R															
R-ASMK⋆ [48]	NC-clean	R50	–	CLS,LOCAL	✓			–	–	–	69.9	**78.7**	45.6	**57.7**	–
R-ASMK⋆+SP [48]	NC-clean	R50	–	CLS,LOCAL	✓			–	–	–	**71.9**	78.0	**48.5**	54.0	–
Global Descriptors															
DIR [47]	SfM-120k	R101	RMAC	TP	✓	–	–	2048	79.0	86.3	53.5	68.3	25.5	42.4	59.2
Radenovic *et al.* [36, 34]	SfM-120k	R101	GeM	SIA		–	–	2048	87.8	**92.7**	64.7	77.2	38.5	56.3	69.5
AGeM [9]	SfM-120k	R101	GeM	SIA		–	–	2048	–	–	**67.0**	**78.1**	**40.7**	**57.3**	–
SOLAR [47]	SfM-120k	R101	GeM	TP,SOS	✓	–	–	2048	78.5	86.3	52.5	70.9	27.1	46.7	60.3
GLAM [46]	SfM-120k	R101	GeM	AF		–	–	512	**89.7**	91.1	66.2	77.5	39.5	54.3	**69.7**
DOLG [47]	SfM-120k	R101	GeM,GAP	AF		–	–	512	72.8	74.5	46.4	56.6	18.1	26.6	49.2
Global Descriptors+D2R															
Mei *et al.* [26]	[O]	R101	FC	CLS				4096	38.4	–	–	–	–	–	–
Salvador *et al.* [43]	Pascal VOC	V16	GSP	CLS,LOCAL		✓		512	67.9	72.9	–	–	–	–	–
Chen *et al.* [4]	OpenImageV4 [17]	R50	MAC	MSE		✓		2048	50.2	65.2	–	–	–	–	–
Liao *et al.* [22]	Oxford,Paris	A,V16	CroW	CLS,LOCAL				768	80.1	90.3	–	–	–	–	–
DIR+RPN [8]	NC-clean	R101	RMAC	TP	✓			2048	**85.2**	**94.0**	–	–	–	–	–
CiDeR (Ours)	SfM-120k	R101	GeM	AF		✓	✓	2048	**89.9**	92.0	**67.3**	**79.4**	**42.4**	57.5	**71.4**
CiDeR-FT (Ours)	SfM-120k	R101	GeM	AF	✓	✓	✓	2048	**92.6**	**95.1**	**76.2**	**84.5**	**58.9**	**68.9**	**79.4**

Performance Comparison

- Superior performance on various datasets

Method	Base		Medium								Hard							
	Ox5k	Par6k	$\mathcal{R}$Oxf		$\mathcal{R}$Oxf +$\mathcal{R}$1M		$\mathcal{R}$Par		$\mathcal{R}$Par +$\mathcal{R}$1M		$\mathcal{R}$Oxf		$\mathcal{R}$Oxf +$\mathcal{R}$1M		$\mathcal{R}$Par		$\mathcal{R}$Par +$\mathcal{R}$1M	
	mAP	mAP	mAP	mP@10	mAP	mP@10	mAP	mP@10	mAP	mP@10	mAP	mP@10	mAP	mP@10	mAP	mP@10	mAP	mP@10
Global Descriptors (SfM-120k)																		
DIR [47]	79.0	86.3	53.5	76.9	–	–	68.3	97.7	–	–	25.5	42.0	–	–	42.4	83.6	–	–
Filip *et al.* [36, 34]	87.8	92.7	64.7	**84.7**	**45.2**	**71.7**	77.2	**98.1**	**52.3**	**95.3**	38.5	**53.0**	**19.9**	**34.9**	56.3	**89.1**	24.7	**73.3**
AGeM [9]	–	–	**67.0**	–	–	–	**78.1**	–	–	–	**40.7**	–	–	–	57.3	–	–	–
SOLAR [47]	78.5	86.3	52.5	73.6	–	–	70.9	**98.1**	–	–	27.1	41.4	–	–	46.7	83.6	–	–
GeM [47]	79.0	82.6	54.0	72.5	–	–	64.3	92.6	–	–	25.8	42.2	–	–	36.6	67.6	–	–
GLAM [47]	**89.7**	91.1	66.2	–	–	–	77.5	–	–	–	39.5	–	–	–	54.3	–	–	–
DOLG [47]	72.8	74.5	46.4	66.8	–	–	56.6	91.1	–	–	18.1	27.9	–	–	26.6	62.6	–	–
CiDeR (Ours)	**89.9**	92.0	**67.3**	**85.1**	**50.3**	**75.5**	**79.4**	97.9	51.4	**95.7**	**42.4**	**56.4**	**22.4**	**35.9**	57.5	87.1	22.4	69.4
CiDeR-FT (Ours)	**92.6**	**95.1**	**76.2**	**87.3**	**60.5**	**78.6**	**84.5**	98.0	**56.9**	**95.9**	**58.9**	**71.1**	**36.8**	**55.7**	**68.9**	**91.3**	**30.1**	**73.9**
Global Descriptors ($\mathcal{R}$GLDv2-clean)																		
Yokoo *et al.* [60]† (Base)	86.1	93.9	64.5	81.0	51.3	72.1	84.1	**95.4**	54.2	90.3	35.6	51.5	22.2	42.9	**68.7**	86.4	27.4	66.9
SOLAR [27]†	90.6	**94.4**	70.8	84.6	55.8	76.1	80.3	94.6	57.6	**92.0**	48.0	**62.3**	30.3	45.3	61.8	83.9	30.7	71.6
GLAM [46]‡	**90.9**	94.1	**72.2**	84.7	**58.6**	76.1	83.0	95.0	**58.6**	91.7	**49.6**	61.6	**34.1**	**50.9**	65.6	**87.6**	**33.3**	72.1
DOLG [59]†	88.3	93.9	70.8	**85.3**	57.3	**76.8**	**83.2**	**95.4**	57.3	**92.0**	47.4	60.0	29.5	46.2	67.9	87.4	32.7	**72.4**
Token [58]†	81.2	89.6	60.8	77.7	44.0	60.9	75.8	94.3	44.1	86.9	37.3	54.1	23.2	37.7	54.8	81.3	19.7	54.4
CiDeR (Ours)	89.8	**94.6**	**73.7**	**85.5**	**58.6**	76.3	**84.6**	**96.7**	**59.0**	**95.1**	**54.9**	**66.6**	**34.6**	**54.7**	**68.5**	**89.1**	**33.5**	**76.9**
CiDeR-FT (Ours)	**90.9**	**96.1**	**77.8**	**88.0**	**61.8**	**78.0**	**87.4**	**97.0**	**61.6**	**94.3**	**61.9**	**70.4**	**39.4**	**56.8**	**75.3**	**90.0**	**35.8**	**72.7**

Conclusion