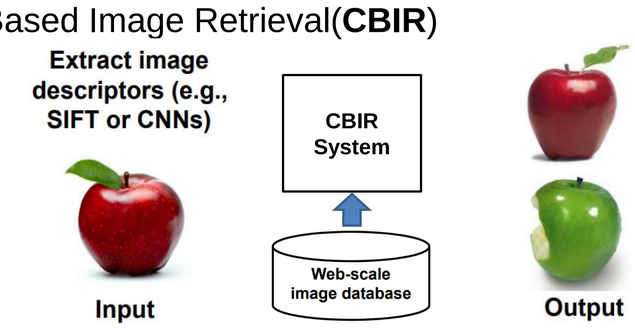
Regional Attention Based Deep Feature for Image Retrieval

Jaeyoon Kim Sung-Eui Yoon
KAIST

Trained model and source codes are available
https://sglab.kaist.ac.kr/RegionalAttention/

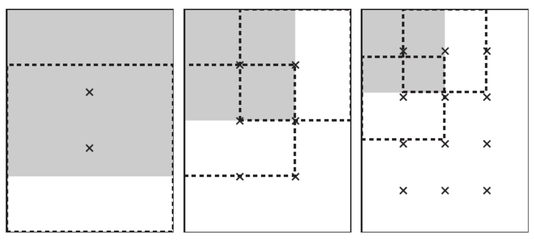
1. Intro. & Motivation

- Content-Based Image Retrieval(**CBIR**)



- Challenge in CBIR – Backgrounds and Clutters

- Regional Maximum Activation of Convolutions(**R-MAC**)[1]
  : R-MAC is sometimes suffered from backgrounds and clutters since it uniformly samples regions of an image.

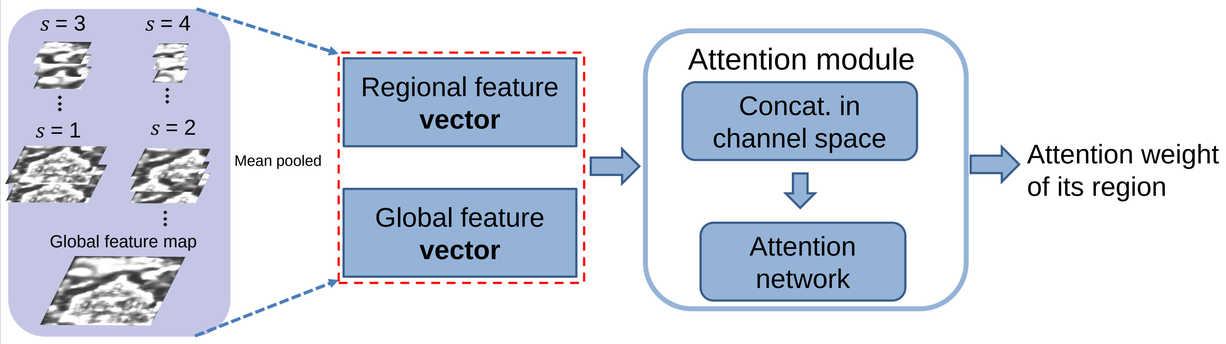


[1] Tolias, G., Sicre, R., Jegou, H.: Particular object retrieval with integral max-pooling of cnn activations. In: ICLR. (2016)

- Categories in Image Retrieval
  - "Fine-tuned" : Fine-tuned **CNN** for specific dataset(data category).
  - "**Pre-trained**": Off-the-shelf **CNN** from ImageNet

3. Context-Aware Regional Attention

- Proposed Regional Attention Network
  - Context awareness: Consider both of **local** and **global** feature of its region.



- Attention Network
  -Two linear layers and Two non-linear layers

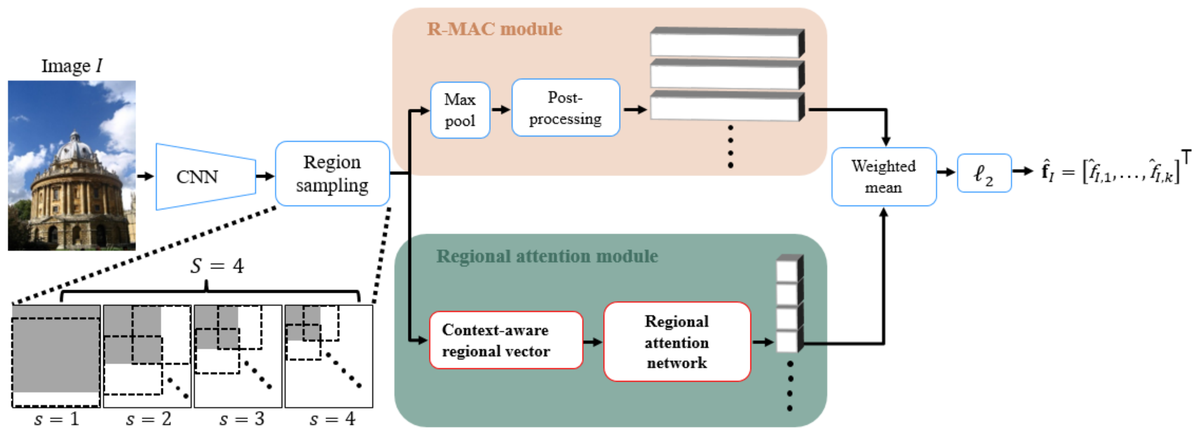
$$\Phi(\mathbf{k}) = \mathrm{softplus}(\mathbf{W_c}\pi(\mathbf{k}) + \mathbf{b_c}),$$
$$\pi(\mathbf{k}) = \tanh(\mathbf{W_r}\mathbf{k} + \mathbf{b_r}).$$

$\mathbf{k}$: combined feature vector(local, global), $\Phi(\mathbf{k})$: Attention weight of $\mathbf{k}$

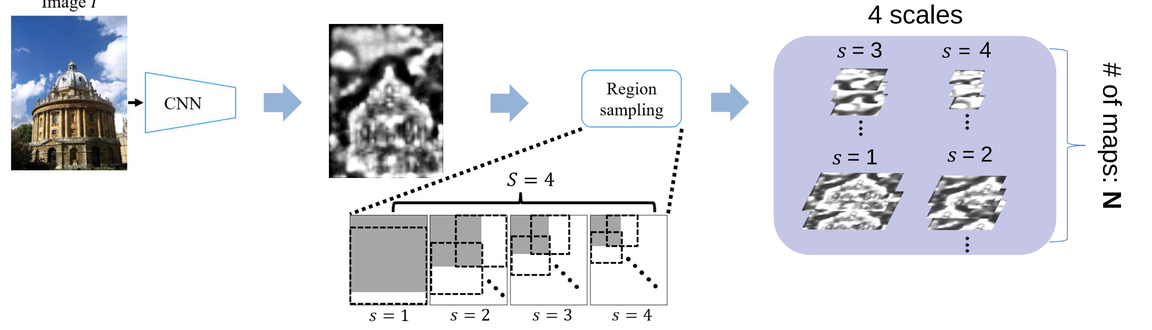
- Training the context-aware regional attention network
  -Parameters to train: $\mathbf{W}_r, b_r, \mathbf{W}_c, b_c$
  -Freezing the CNN(Resnet101) while training our attention network
  -Dataset: ILSVRC2012-ImageNet for "**Pre-trained**" category
  -Classification loss(Cross entropy)

2. Image Encoding Pipeline

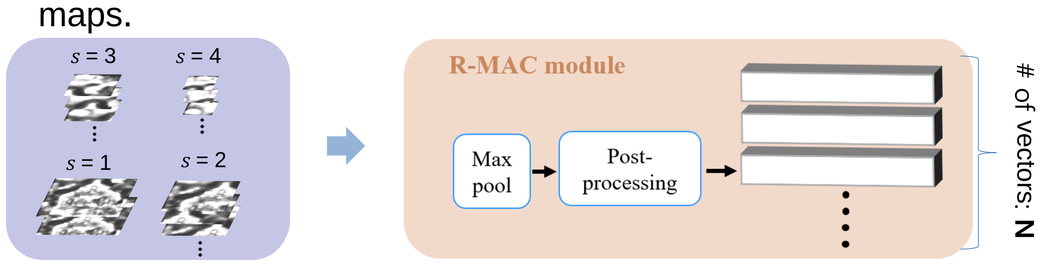
- Overall Sequence of encoding pipeline



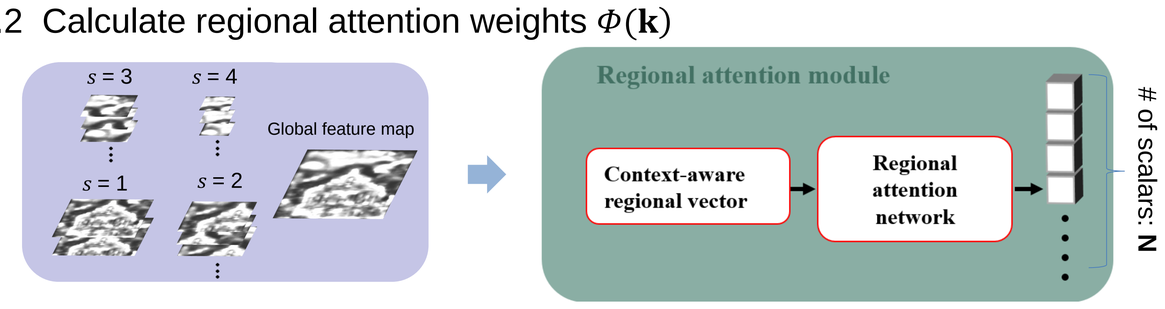
1. Extract a CNN feature map and sample **regional feature maps** in a R-MAC manner



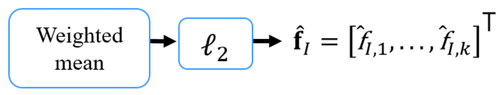
2.1 Produce **R-MAC feature** vectors with the regional feature maps.



2.2 Calculate regional attention weights $\Phi(\mathbf{k})$



3. Obtain a global feature vector, $\hat{\mathbf{f}}_I$, through combining R-MAC features with regional attention weights



4. Experiments

Method	Scale (S)			
	S=3	S=4	S=5	S=6
Baseline	69.9	**70.7**	70.1	69.0
Ours	75.1	76.7	**76.8**	76.4

(a)

Method	Oxford5k	Paris6k
RPN + PCA Landmark	64.7	75.5
+ Regional attention	66.6	75.8
+ Context awareness	67.9	76.4

(b)

Ablation study

Method	Oxford5k	Paris6k	Time (s)
Baseline + PCA Landmark	70.1	85.4	0.095
+ Regional attention	74.9	86.0	0.115
+ Context awareness	76.8	87.5	0.123
- PCA Landmark + PCA Paris, Oxford	77.6	88.3	-

(c)

Comparison with state-of-the-arts

	Method	Dim.	Oxford5k	Paris6k	Oxford105k	Paris106k
Resnet101	SDCF [9]	2048	69.1	81.7	65.4	74.3
	CroW [14]	2048	68.7	82.8	62.7	75.1
	R-MAC [27]	2048	70.1	85.4	66.9	80.8
	CAM [1]	2048	69.9	84.3	64.3	77.1
	Ours	2048	**76.8**	**87.5**	**73.6**	**82.5**
	Query expansion (QE)					
Resnet101	SDCF+QE [9]	2048	68.5	84.9	66.8	79.4
	CroW+QE [14]	2048	69.5	85.1	66.7	79.9
	R-MAC+QE [27]	2048	73.8	86.4	71.8	82.6
	CAM+QE [1]	2048	71.3	86.1	68.7	80.8
	Ours+QE	2048	**81.8**	**89.3**	**80.4**	**85.4**

BMVC 2018
Newcastle upon Tyne

Trained model and source codes are available
https://sglab.kaist.ac.kr/RegionalAttention/

KAIST