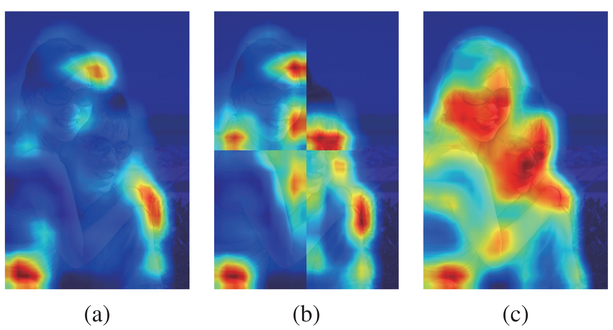
Paper ID : 2479

PUZZLE-CAM: IMPROVED LOCALIZATION VIA MATCHING PARTIAL AND FULL FEATURES

Sanghyun Jo — GYNetworks
*In-Jae Yu** — KAIST School of Computing

IEEE Signal Processing Society
ICIP2021 Anchorage, Alaska.

Motivation



(a) (b) (c)

Fig 1. Illustration of the motivation of this work.

Observation

- The conventional CAMs activates part of objects for classification (a)
- The merged CAMs from image patches scatter the attention (b)
- Puzzle-CAM suppresses the attention on discriminative region of the object (c)

Process

1. Tilling an image to image patches to divide into the attention.
2. Merging the feature maps from the network to produce the reconstructed features.
3. Matching partial and full features with reconstructing regularization.

Puzzle-CAM

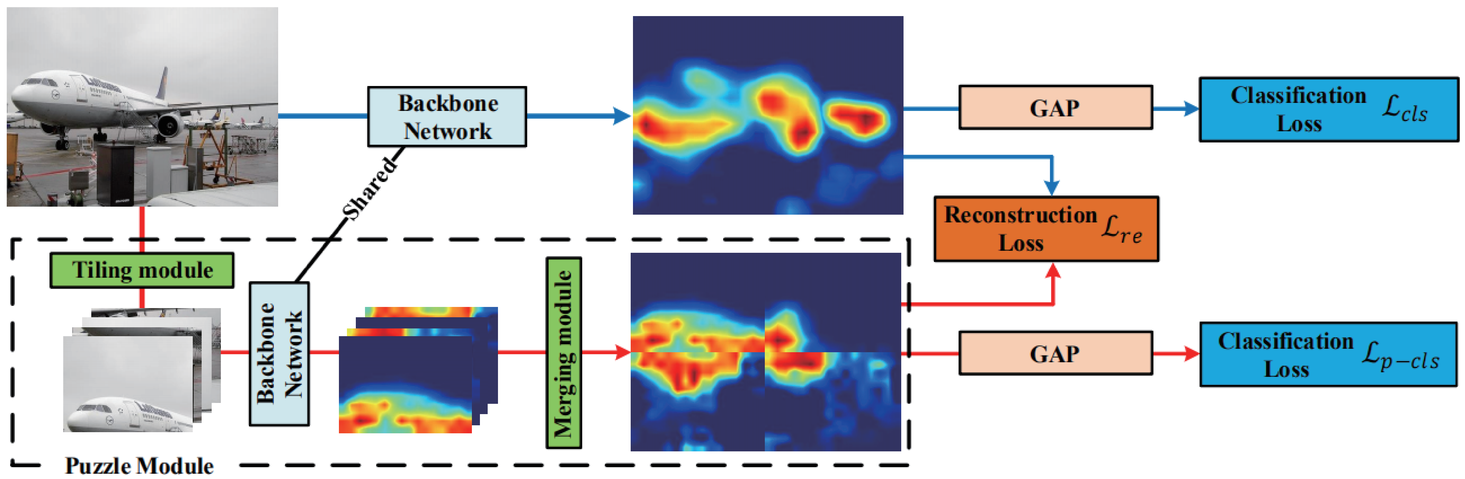


Fig 2. Illustration of the Puzzle-CAM architecture.

Proposed Losses

$$\mathcal{L}_{cls} = \ell_{cls}(\hat{Y}^s, Y)$$
$$\mathcal{L}_{p-cls} = \ell_{cls}(\hat{Y}^{re}, Y)$$
$$\mathcal{L}_{re} = \|A^s - A^{re}\|_1$$
$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{p-cls} + \alpha\mathcal{L}_{re}$$

✔ Ablation study: using different loss functions

L_{cls}	L_{p-cls}	L_{re}	mIoU (%)
✓			47.82
✓	✓		47.70
✓		✓	49.21
✓	✓	✓	**51.53**

Quantitative Results

✔ Ablation study: using different backbones

Method	Backbone	CAM (%)	CAM +RW (%)	CAM+RW +dCRF (%)
AffinityNet [4]	ResNet-50	47.82	58.10	59.70
Puzzle-CAM	ResNet-50	51.53	64.16	64.70
Puzzle-CAM	ResNeSt-50	57.59	69.48	69.91
Puzzle-CAM	ResNeSt-101	61.85	71.92	72.46
Puzzle-CAM	ResNeSt-269	**62.45**	**74.14**	**74.67**

✔ Comparison with existing state-of-the-art methods

Method	Backbone	Supervision	val	test
AffinityNet [4]	Wide-ResNet-38	$\mathcal{I}$	61.7	63.7
DSRG [12]	ResNet-101	$\mathcal{I}+\mathcal{S}$	61.4	63.2
SeeNet [13]	ResNet-101	$\mathcal{I}+\mathcal{S}$	63.1	62.8
IRNet [4]	ResNet-50	$\mathcal{I}$	63.5	64.8
FickleNet [6]	ResNet-101	$\mathcal{I}+\mathcal{S}$	64.9	65.3
ICD [17]	ResNet-101	$\mathcal{I}$	64.1	64.3
SEAM [5]	Wide-ResNet-38	$\mathcal{I}$	64.5	65.7
Ours (Puzzle-CAM)	ResNeSt-101	$\mathcal{I}$	66.9	67.7
Ours (Puzzle-CAM)	ResNeSt-269	$\mathcal{I}$	**71.9**	**72.2**

Fig 3. (a): original images. (b): ground truth. (c): segmentation results predicted by AffinityNet. (d): segmentation results predicted by Puzzle-CAM.