SaliencyMix: A Saliency Guided Data Augmentation Strategy for Better Regularization

A F M Shahab Uddin, Mst Sirazam Monira, Wheemyung Shin, TaeChoong Chung and Sung-Ho Bae

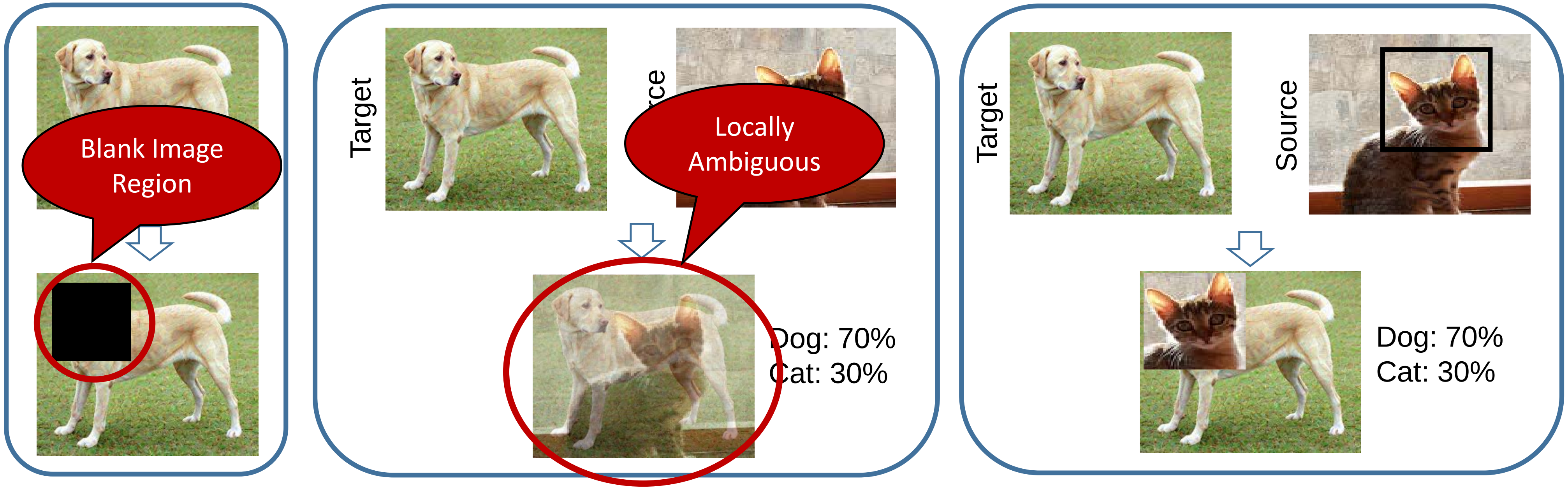
Department of Computer Science and Engineering

Kyung Hee University, South Korea.

ICLR 2021

Advanced Data Augmentation Techniques

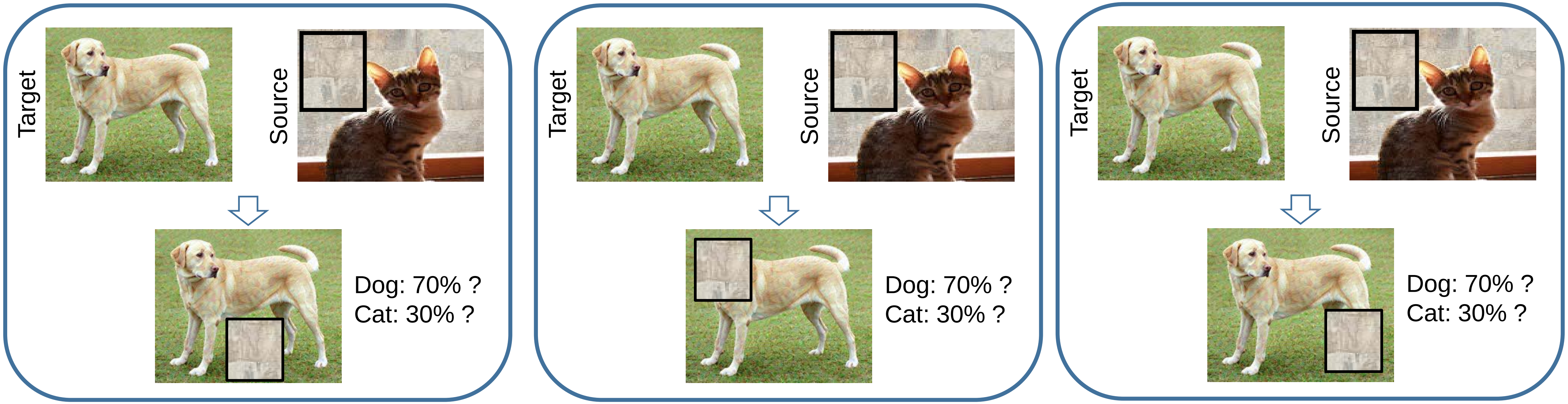
- Better **regularization** technique helps to **improve** model **robustness** and **performance**
- **Recently**, several **effective data augmentation** strategies have been proposed



Cutout [1]   Mixup [2]   CutMix [3]

Motivation

- Limitation of CutMix [3]
  - o **Random** selection of the **source patch** may **not** always **represent** the source object

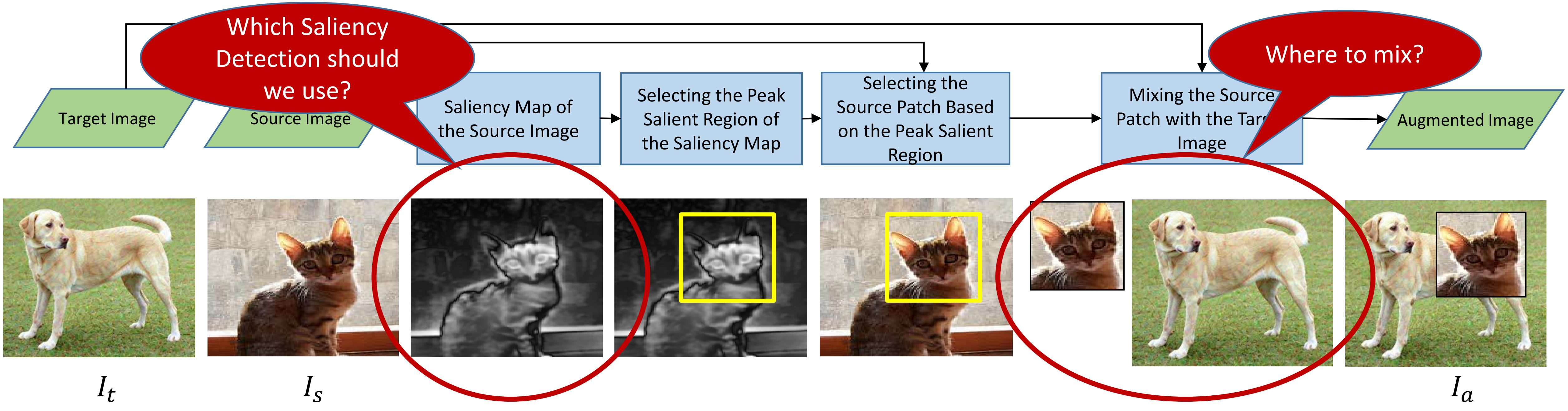


Guides the model to learn  as Cat feature !!!

May **mislead** the classifier?

Our Approach: SaliencyMix

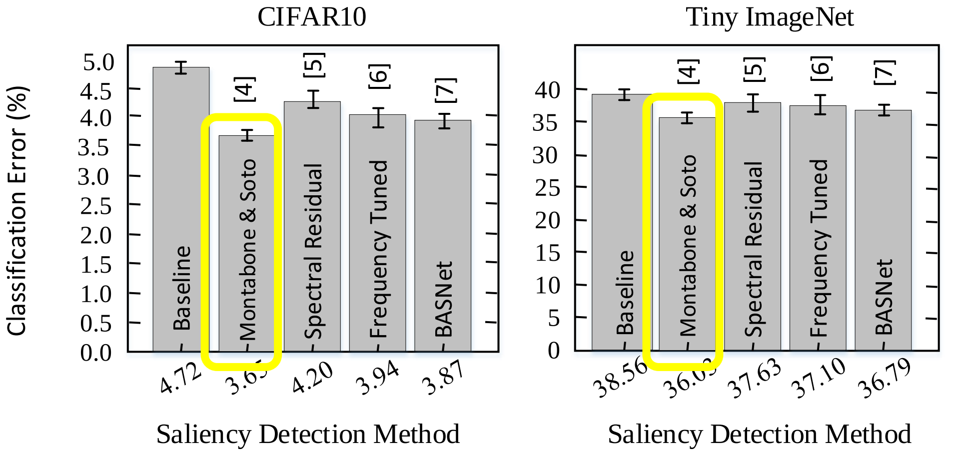
- Proposed approach
  1. Extract the **Saliency map** [4] of the source image
  2. Select and cut the **most salient region** of the source image
  3. Then **mix** the source patch with the **target image** based on a **mixing ration λ**



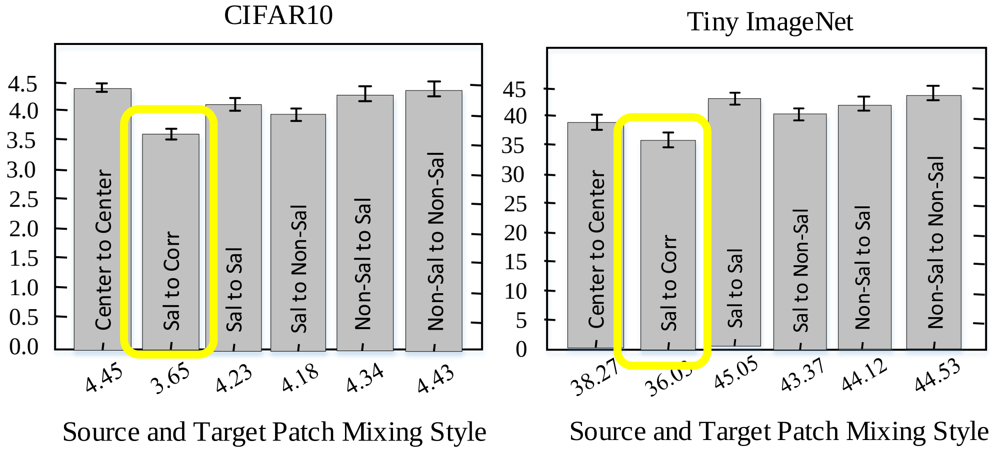
Augmented Image $I_a = M \odot I_s + M' \odot I_t$

Augmented Label $y_a = \lambda y_t + (1 - \lambda) y_s$

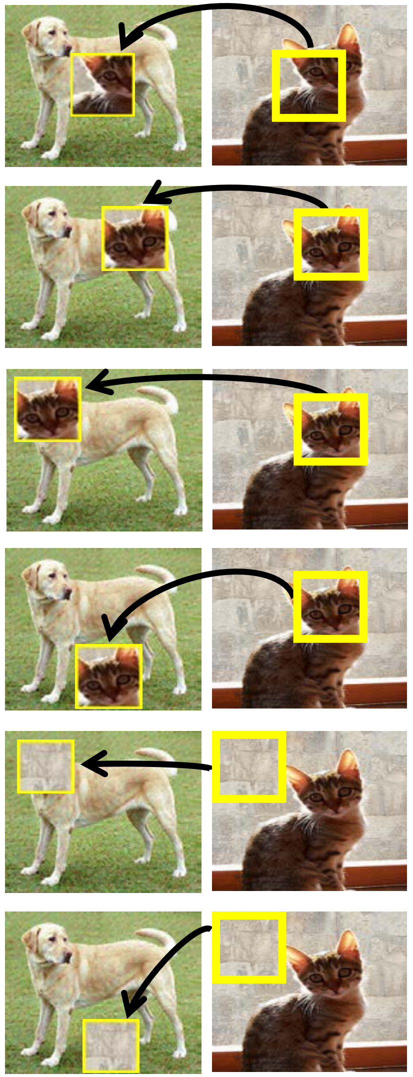
- Effect of **various Saliency Detection** methods



- Several strategy of **source patch selection and mixing**



(c) (d)



Experimental Result

Image Classification

❖ **CIFAR** dataset

Method	Top-1 Error (%) CIFAR-10	CIFAR-10+	CIFAR-100	CIFAR-100+
ResNet-18 (Baseline)	10.63± 0.26	4.72± 0.21	36.68± 0.57	22.46± 0.31
ResNet-18 + Cutout	9.31±0.18	3.99±0.13	34.98±0.29	21.96±0.24
ResNet-18 + CutMix	9.44±0.34	3.78±0.12	34.42±0.27	19.42±0.23
ResNet-18 + SaliencyMix	**7.59±0.22**	**3.65±0.10**	**28.73±0.13**	**19.29±0.21**
ResNet-50 (Baseline)	12.14±0.95	4.98±0.14	36.48±0.50	21.58±0.43
ResNet-50 + Cutout	8.84±0.77	3.86±0.25	32.97±0.74	21.38±0.69
ResNet-50 + CutMix	9.16±0.38	3.61±0.13	31.65±0.61	18.72±0.23
ResNet-50 + SaliencyMix	**6.81±0.30**	**3.46±0.08**	**24.89±0.39**	**18.57±0.29**
WideResNet-28-10 (Baseline)	6.97±0.22	3.87±0.08	26.06±0.22	18.80±0.08
WideResNet-28-10 + Cutout	5.54±0.08	3.08±0.16	23.94±0.15	18.41±0.27
WideResNet-28-10 + AutoAugment	-	**2.60±0.10**	-	17.10±0.30
WideResNet-28-10 + PuzzleMix (200 epochs)	-	-	-	**16.23**
WideResNet-28-10 + CutMix	5.18±0.20	2.87±0.16	23.21±0.20	16.66±0.20
WideResNet-28-10 + SaliencyMix	**4.04±0.13**	2.76±0.07	**19.45±0.32**	16.56±0.17

❖ **ImageNet** dataset

Method	Top-1 Error (%)	Top-5 Error (%)
ResNet-50 (Baseline)	23.68	7.05
ResNet-50 + Cutout	22.93	6.66
ResNet-50 + StochasticDepth	22.46	6.27
ResNet-50 + Mixup	22.58	6.40
ResNet-50 + Manifold Mixup	22.50	6.21
ResNet-50 + AutoAugment	22.40	6.20
ResNet-50 + DropBlock	21.87	5.98
ResNet-50 + CutMix	21.40	5.92
ResNet-50 + PuzzleMix	**21.24**	**5.71**
ResNet-50 + SaliencyMix	**21.26**	**5.76**
ResNet-101 (Baseline)	21.87	6.29
ResNet-101 + Cutout	20.72	5.51
ResNet-101 + Mixup	20.52	5.28
ResNet-101 + CutMix	20.17	5.24
ResNet-101 + SaliencyMix	**20.09**	**5.15**

Experimental Result

❖ **Transfer Learning on Object Detection Task**

Backbone Network	ImageNet Cls. Err. Top-1 (%)	Detection (F-RCNN) (mAP)
ResNet-50 (Baseline)	23.68	76.71 (+0.00)
Cutout-trained	22.93	77.17 (+0.46)
Mixup-trained	22.58	77.98 (+1.27)
CutMix-trained	21.40	78.31 (+1.60)
SaliencyMix-trained	**21.26**	**78.48 (+1.77)**

👉 SaliencyMix trained model offers **1.77 %** performance **improvement**

❖ **Adversarial Robustness**

	Baseline	Cutout	Mixup	CutMix	SaliencyMix
Acc. (%)	8.2	11.5	24.4	31.0	**32.96**

👉 SaliencyMix trained model shows **better robustness**

❖ **Computational Complexity**

	Baseline	Cutout	Mixup	CutMix	SaliencyMix
Time (Hour)	**0.83**	0.84	0.87	0.89	0.91

👉 Computational burden slightly **increased** due to saliency detection

Conclusion

- **SaliencyMix** is an **effective data augmentation** technique
  - Offers the CNN with **greater regularization** ability
  - **Improves** the model **performance**
    - *Classification*
      - 👉 WideResNet: Best known top-1 error of **2:76%** and **16:56%** on CIFAR-10 and CIFAR-100, respectively
      - 👉 ResNet-50: Best known top-1 error of **21.26%** on ImageNet
      - 👉 ResNet-101: Best known top-1 error of **20.09%** on ImageNet
    - *Object Detection*
      - 👉 SaliencyMix trained model improves detection performance by **+1.77** mAP
    - *Robustness against adversarial attack*
      - 👉 SaliencyMix trained model achieves **1:96%** accuracy improvement on adversarially perturbed ImageNet validation set

Thank you

References

[1] T. DeVries and G. W. Taylor. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552, 2017.

[2] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412, 2017.

[3] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. ICCV, 2019.

[4] S. Montabone and A. Soto, "Human detection using a mobile platform and novel features derived from a visual saliency mechanism," Image and Vision Computing, vol. 28, no. 3, pp. 391-402, 2010.

[5] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in 2007 IEEE Conference on Computer Vision and Pattern Recognition, June 2007, pp. 1-8.

[6] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, June 2009, pp. 1597-1604.

[7] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "Basnet: Boundary-aware salient object detection," in The IEEE Conference on Computer Vision and Pattern Recognition, June 2019.