HITACHI Inspire the Next

Spatial Sensitive Grad-CAM++: Improved Visual Explanation for Object Detectors via Weighted Combination of Gradient Map

Toshinori Yamauchi   Hitachi, Ltd. Research & Development Group

CVPR JUNE 17-21, 2024 SEATTLE, WA
The 3rd XAI4CV Workshop

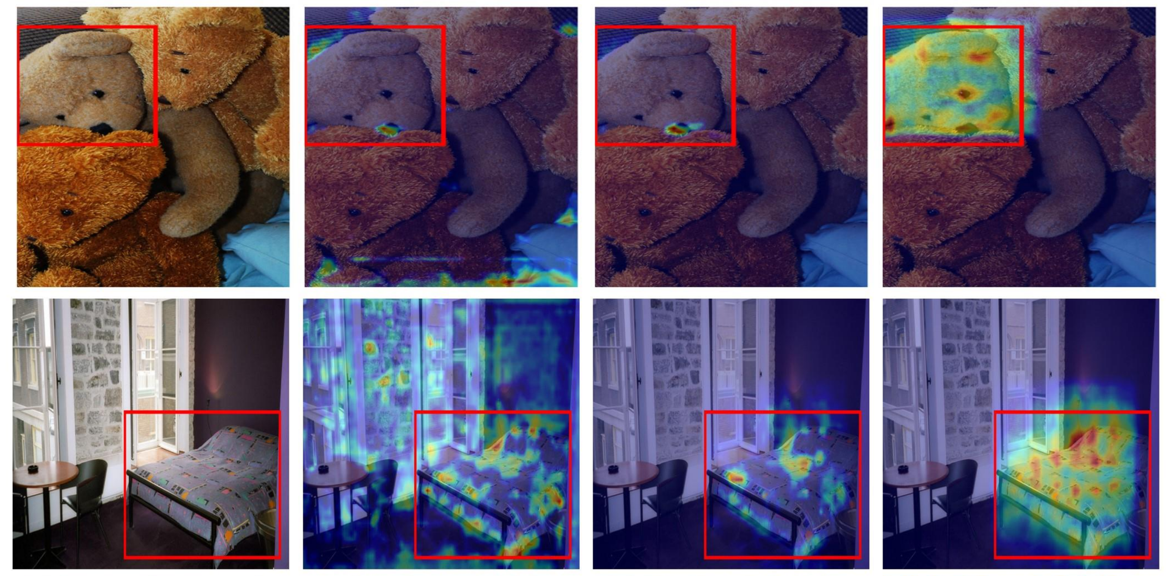
1. Introduction

Visual explanations for object detection

- Object detectors detect each instance in an image.
- Instance-specific heat maps need to be generated for each detected instance.
- Generated heat maps by existing methods may only highlight a part of important regions.

Contribution

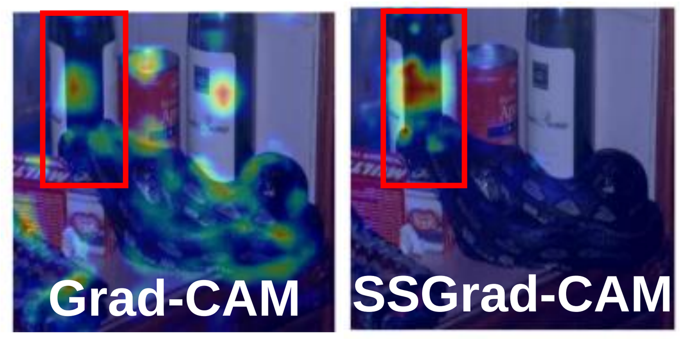
- Propose SSGrad-CAM++, which improves SSGrad-CAM by incorporating weighted combination of gradient maps.
- SSGrad-CAM++ can generate more reliable heat maps.



(a) Detection result (b) Grad-CAM (c) SSGrad-CAM (d) SSGrad-CAM++

2. Base Method: Spatial Sensitive Grad-CAM

- It computes the importance of both features and space to generate instance-specific heat maps.



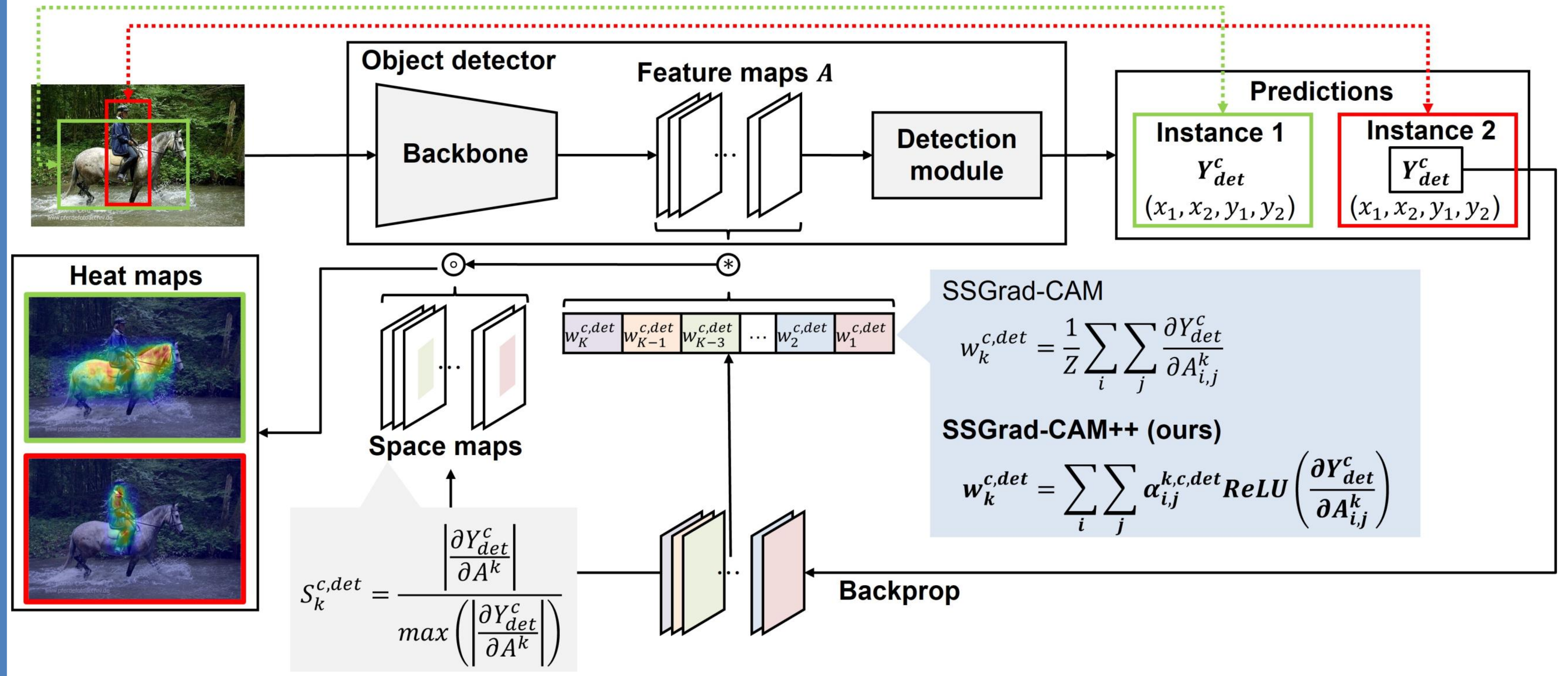
$$L^{c,det} = ReLU\left(\sum_k \overbrace{\left(w_k^{c,det} A^k\right)}^{\text{Grad-CAM}} \odot S_k^{c,det}\right)$$

Y^c_{det}: Predicted score
A^k : Feature map

$$w_k^{c,det} = \frac{1}{Z}\sum_i\sum_j \frac{\partial Y^c_{det}}{\partial A^k_{i,j}} \qquad S_k^{c,det} = \frac{\left|\frac{\partial Y^c_{det}}{\partial A^k}\right|}{\max\left(\left|\frac{\partial Y^c_{det}}{\partial A^k}\right|\right)}$$

Importance of features   Importance of space

3. Proposed Method: Spatial Sensitive Grad-CAM++



Room for improvement in computation of the importance of features

- SSGrad-CAM computes the importance of features by averaging the gradient map.
- The magnitude of $w_k^{c,det}$ depends on the spatial size of the feature in A^k.

$$w_k^{c,det} = \frac{1}{Z}\sum_i\sum_j \frac{\partial Y^c_{det}}{\partial A^k_{i,j}}$$

Importance of features

Incorporation of weighted combination of gradient maps

Definition of the importance of features

$$w_k^{c,det} = \sum_i\sum_j \alpha_{i,j}^{k,c,det} ReLU\left(\frac{\partial Y^c_{det}}{\partial A^k_{i,j}}\right)$$

Assumption for score.

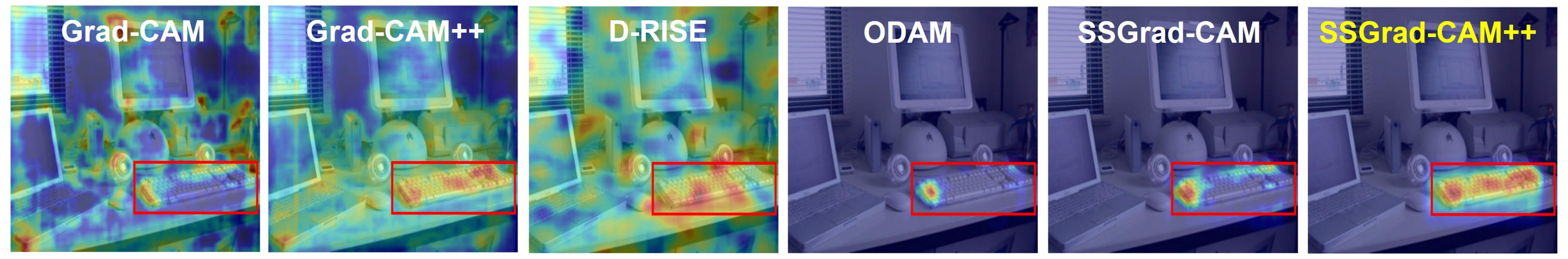
$$Y^c_{det} = \sum_k w_k^{c,det} \sum_i\sum_j A^k_{i,j} M^{k,det}_{i,j}$$

$$\alpha_{i,j}^{k,c,det} = \frac{\frac{\partial^2 Y^c_{det}}{\left(\partial A^k_{i,j}\right)^2}}{2\frac{\partial^2 Y^c_{det}}{\left(\partial A^k_{i,j}\right)^2} M^{k,det}_{i,j} + \sum_a\sum_b A^k_{a,b} M^{k,det}_{a,b}\left(\frac{\partial^3 Y^c_{det}}{\left(\partial A^k_{i,j}\right)^3}\right)}$$

Calculation of the second derivative

$M^{k,det}_{i,j}$: A binary mask with 1 for regions related to the detected instance

Heat map comparison



3. Experiment

Deletion (Del) and Insertion (Ins)

	$score = s_1$		$score = s_2$	
Method	Del ↓	Ins ↑	Del ↓	Ins ↑
Grad-CAM [10]	0.201	0.650	0.241	0.655
Grad-CAM++ [2]	0.104	0.853	0.142	0.851
D-RISE [8]	0.154	0.781	0.201	0.766
ODAM [18]	0.113	0.745	0.180	0.731
SSGrad-CAM [15]	0.071	0.916	0.135	0.867
ours	**0.055**	**0.942**	**0.102**	**0.901**

Pointing Game

Method	P(b)	P(m)	eP(b)	eP(m)
Grad-CAM [10]	0.389	0.330	0.127	0.08
Grad-CAM++ [2]	0.649	0.563	0.150	0.09
D-RISE [8]	0.624	0.505	0.119	0.070
ODAM [18]	0.916	0.804	0.737	**0.546**
SSGrad-CAM [15]	0.911	0.769	0.726	0.509
ours	**0.981**	**0.880**	**0.743**	0.512

SSGrad-CAM++ outperforms other methods, indicating that it generates higher quality heat maps.

3. Conclusion

- We propose SSGrad-CAM++, CAM-based visual explanations for object detectors.
- SSGrad-CAM++ improves SSGrad-CAM in terms of the computation of the importance of features.
- Our experiments show that SSGrad-CAM++ generates heat maps that more accurately capture the important regions.

References

1. Selvaraju, R. R., et al. 2. Chattopadhay, A., et al. 3. Petsiuk, V., et al.
4. Yamauchi, T. and Ishikawa, M. 5. Zhao, C. and Chan,