Explaining Object Detection Through Difference Map

Shujun Xia, Chenyang Zhao, Antoni B. Chan
Department of Computer Science, City University of Hong Kong

香港城市大學
City University of Hong Kong

ICCV OCT 19-23, 2025 HONOLULU HAWAII

Highlights

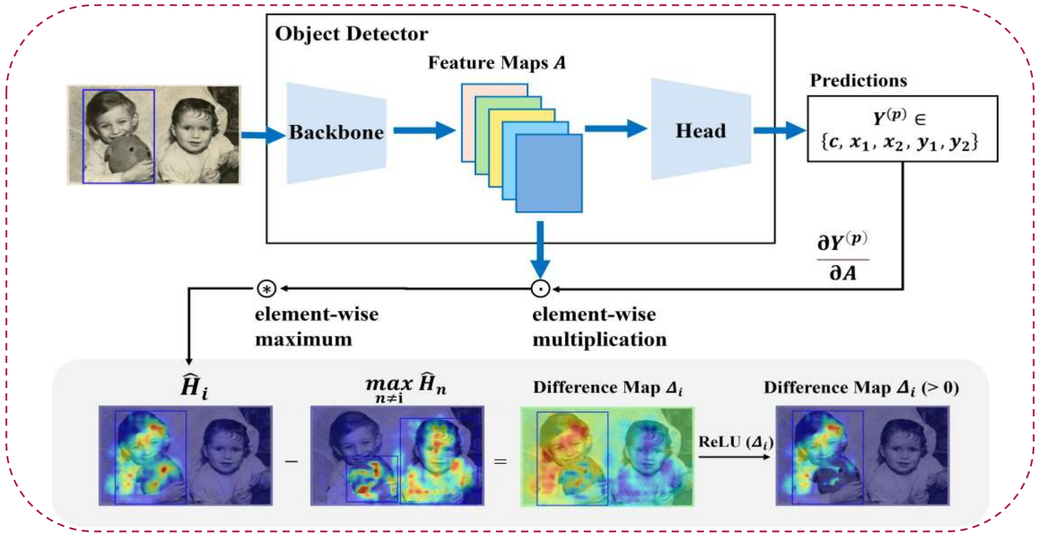
- Enhance the explainability of object detectors by considering **Object Discrimination**, and propose the **difference map** to isolate the key features for a specific object instance's prediction from other objects within the same scene.
- Demonstrate the **generalizability** of the difference map by exhibiting explanations on one- and two-stage, and transformer-based detectors with different types of backbones and detector heads.
- Integrate the difference map with **multiple heatmap-based visual explanation methods** and show that it **improves** the **original heatmaps** by enhancing focus and reducing irrelevant highlights.

Motivation

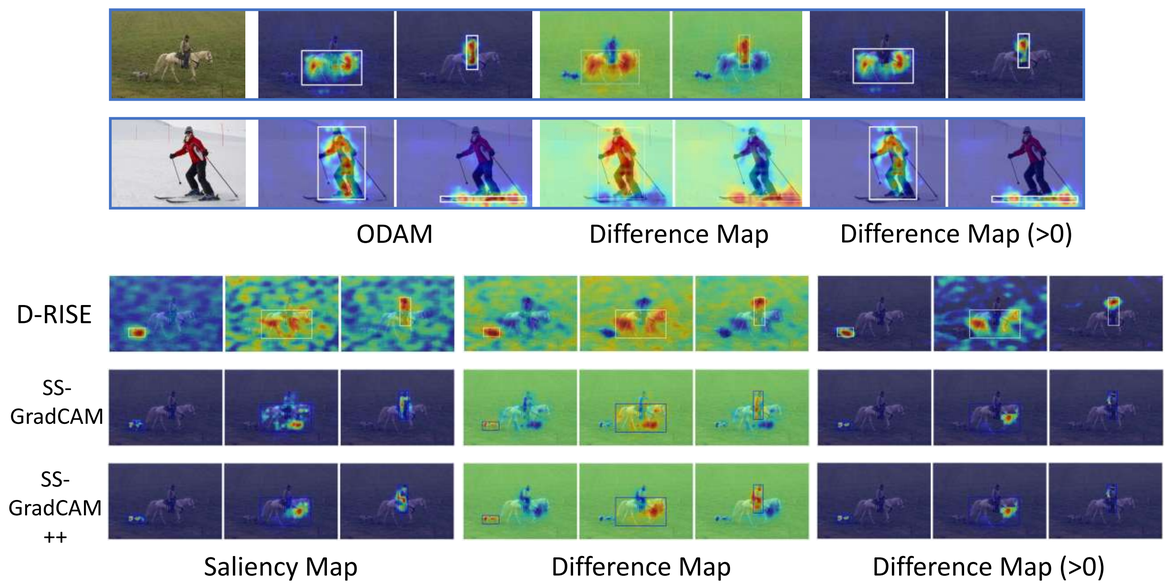
Visual explanation: The explanation approaches produce heat maps locating the regions in the input images that the model looked at, and representing the influence of different pixels on the model's decision.

For object detectors, **instance specific explanations** effectively highlight important regions for each prediction. However, existing approaches have largely **overlooked inter-object relationships** — particularly the relative importance of each pixel across different objects, a concept we refer to as **Object Discrimination (OD)**.

Difference Map Overview

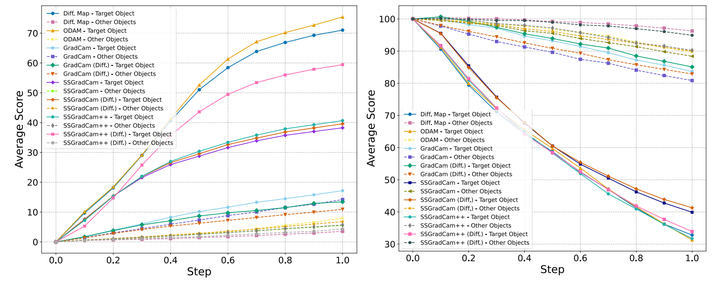


Visualization



Augmenting visual explanation methods with the difference map refines the original saliency maps by effectively suppressing the highlighting of irrelevant or misleading regions associated with detected objects.

Quantitative Evaluation



(a) Insertion Steps (b) Deletion Steps

Method	Target Object Del(↓)	Target Object Ins(↑)	Other Objects Del(↑)	Other Objects Ins(↓)
Grad-CAM	94.88	9.40	89.54	7.23
Grad-CAM (Diff.)	93.76	7.84	91.01	5.99
SS-GradCAM	64.91	25.67	94.99	2.66
SS-GradCAM (Diff.)	65.24	26.33	95.70	3.10
SS-GradCAM++	60.84	26.86	96.27	2.68
SS-GradCAM++ (Diff.)	61.40	37.16	98.44	1.94
ODAM	61.27	**46.06**	96.12	3.14
Difference Map (Ours)	**60.74**	44.25	**99.01**	**1.58**

Table 1: AUC for the Deletion and Insertion curves. In the Deletion process, a lower AUC for the target means removing key pixels greatly reduces its confidence, while a higher AUC for others shows their predictions stay stable. Conversely, in the Insertion process, a higher AUC for the target and lower AUC for others indicate accurate localization of discriminative regions. **Deleting or inserting pixels attributed to the target object should not affect the predictions of other objects in the image.**