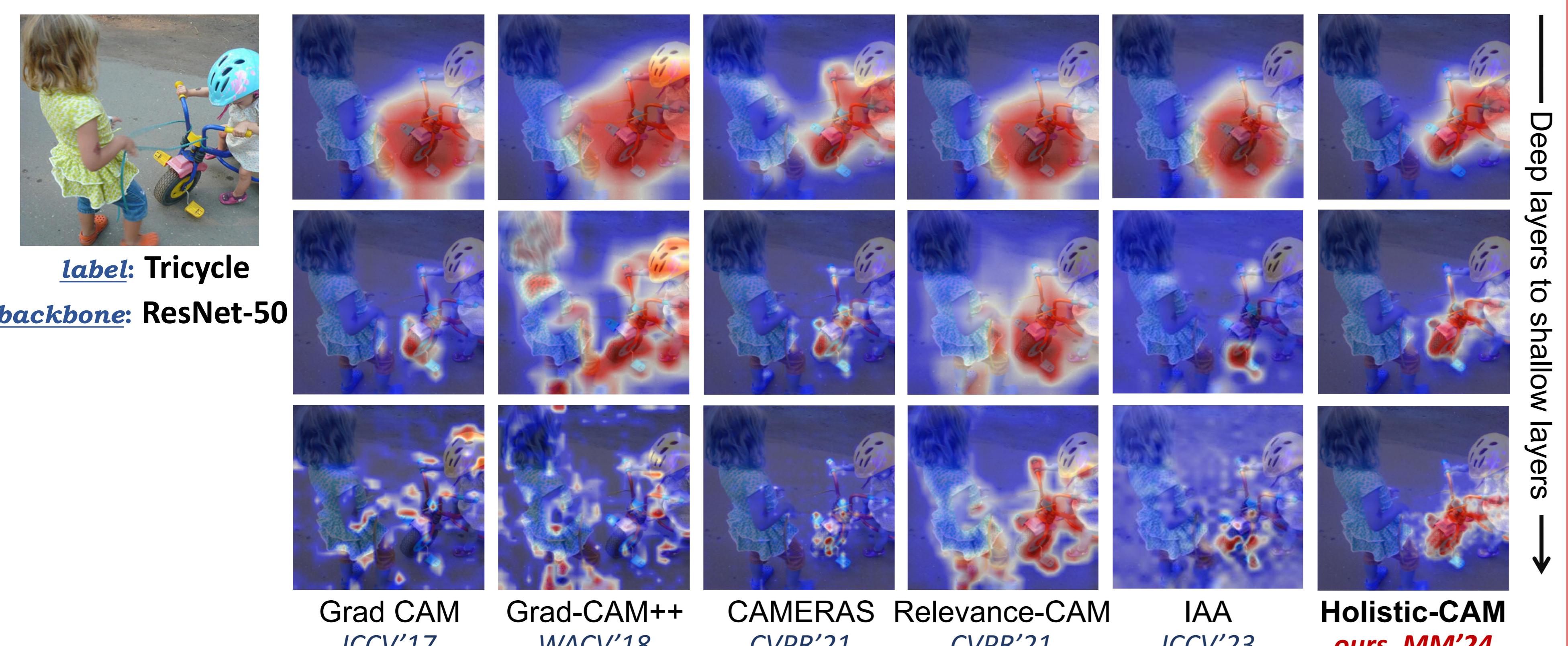


# Holistic-CAM: Ultra-lucid and Sanity Preserving Visual Interpretation in Holistic Stage of CNNs

Pengxu Chen<sup>1</sup>, Huazhong Liu<sup>1,\*</sup>, Jihong Ding<sup>1,\*</sup>, Jiawen Luo<sup>1</sup>, Peng Tan<sup>3</sup>, Laurence T. Yang<sup>2,4</sup>

<sup>1</sup> Hainan University, <sup>2</sup> Zhengzhou University, <sup>3</sup> Hainan Credit Investigation Co. Ltd., <sup>4</sup> St.Francis Xavier University

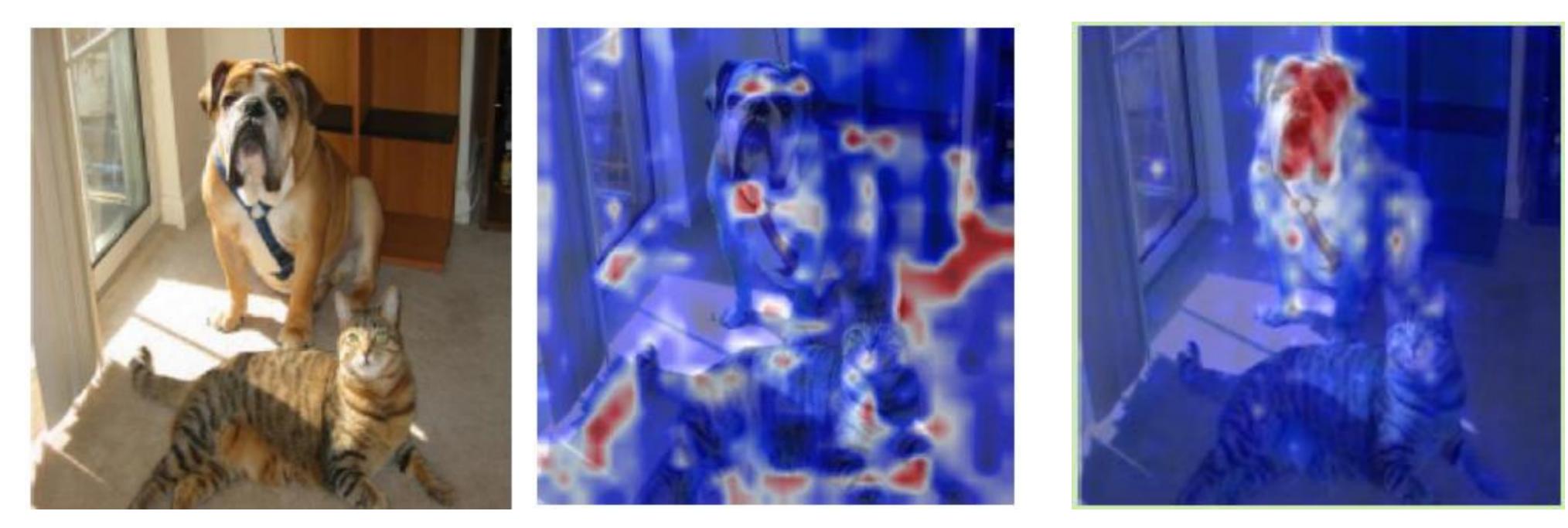
**GOAL:** Interpreting the holistic-stage of media-related deep learning model in a lucid and understandable manner.



**CHALLENGES and CONTRIBUTIONS**

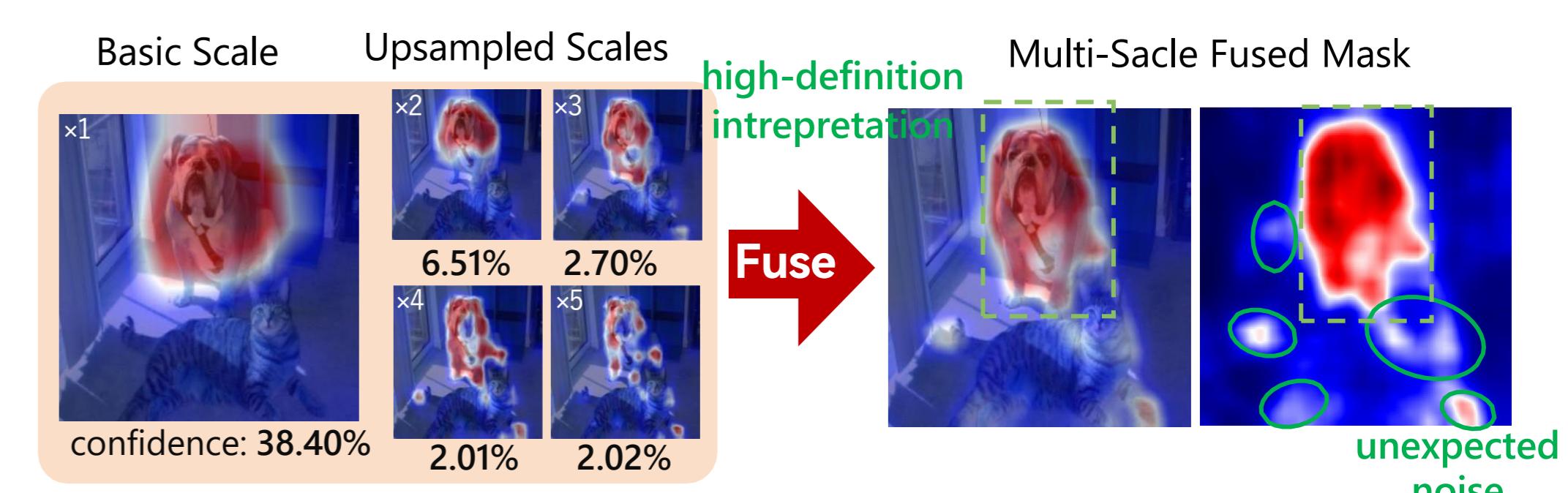
1. Robustless attributions in shallow layers.

**Elem-wise weighted positive gradient (PGE).**

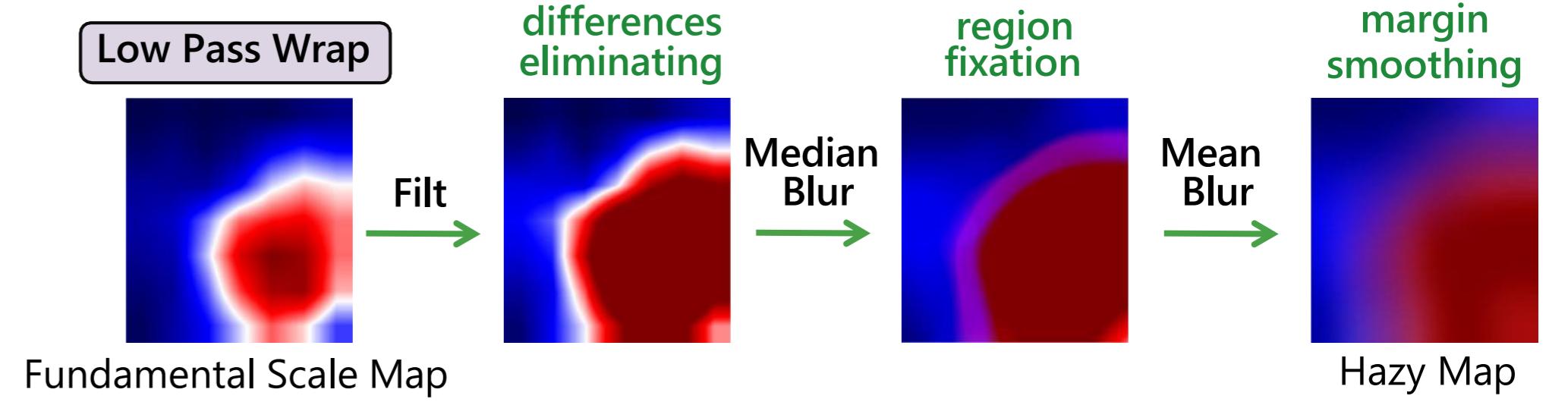


$$PGE(G_{ij}^{kc}) = \alpha_{ij}^{kc} \cdot \text{ReLU}(G_{ij}^{kc}) \\ = \frac{\frac{\partial^2 f_c(x)}{(\partial A_{ij}^k)^2}}{2 \frac{\partial^2 f_c(x)}{\partial (A_{ij}^k)^2} + \frac{\partial^3 f_c(x)}{(\partial A_{ij}^k)^3} \sum_a \sum_b A_{ab}^k + \text{eps}} \cdot \text{relu}\left(\frac{\partial f_c(x)}{\partial A_{ij}^k}\right).$$

2. Limited resolution in deep layers, as well as unfaithful attributions of multi-scale fusion.



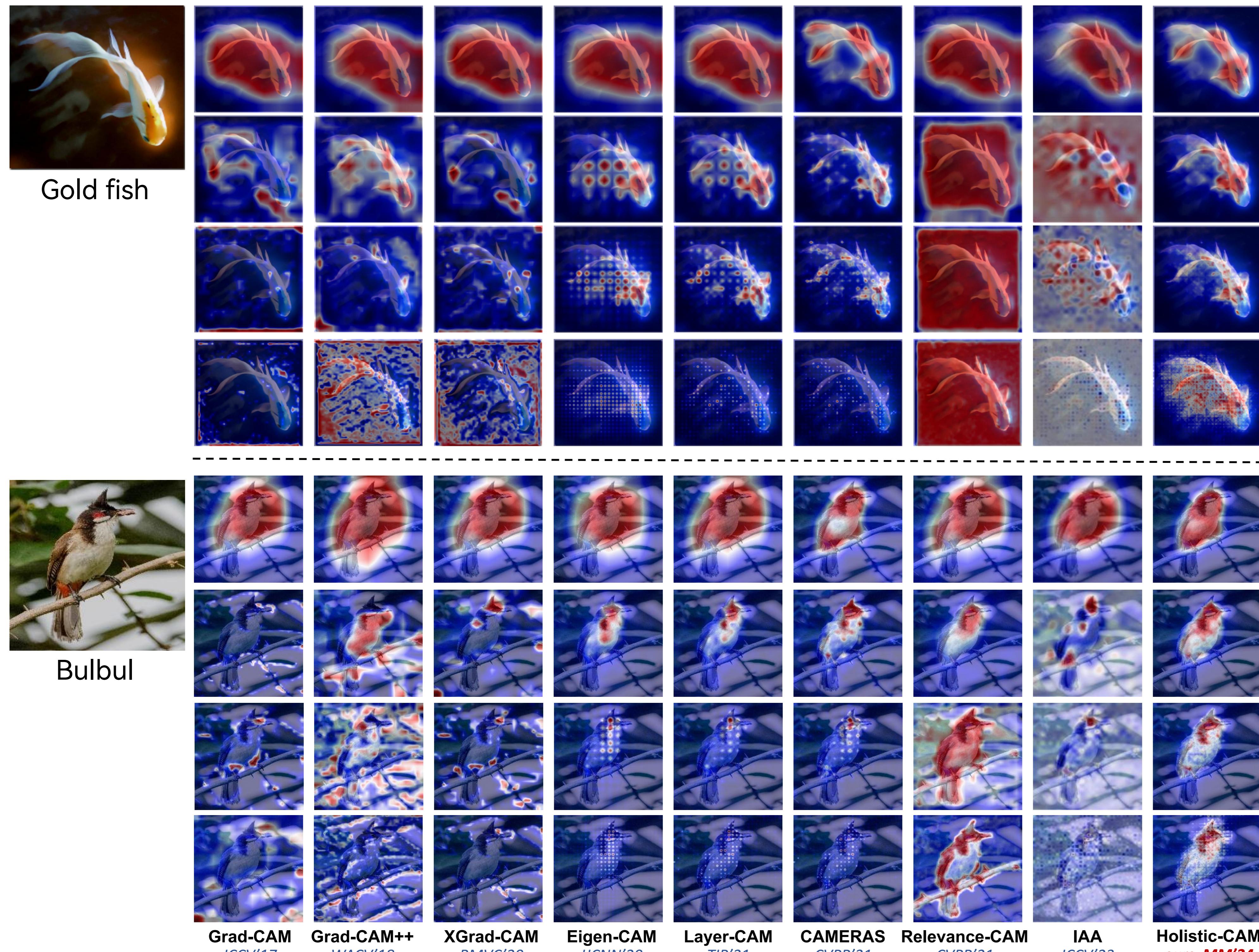
**Multi-scale fusion scheme** with proposed **denoising module** based on fundamental-scale feature and **low-pass wrap** technology.



**Low-pass wrap:** eliminating noises beyond the localization area without compromising the fine-grained details of larger-scale features.

## EVALUATIONS

### Visual Assessments



### Quantitative Experiments

ResNet-50	Ins↑	Del↓	Over-all↑	ROAD↑	EPG↑
Deep Layer [Layer4]	Grad-CAM	54.887	11.622	43.265	28.095
	Grad-CAM++	51.165	14.172	36.993	22.470
	XGrad-CAM	54.887	11.622	43.265	28.099
	Eigen-CAM	53.249	12.705	40.544	25.595
	Layer-CAM	54.018	11.882	42.136	26.799
	CAMERAS	54.439	8.698	45.741	28.606
	Relevance-CAM	54.663	11.622	43.041	27.981
	IAA	54.872	11.901	42.971	28.871
	<b>Holistic-CAM</b>	<b>55.056</b>	<b>8.947</b>	<b>46.109</b>	<b>29.047</b>
	Grad-CAM	18.876	15.207	3.669	4.673

VGG-16	Ins↑	Del↓	Over-all↑	ROAD↑	EPG↑
Deep Layer [Layer43]	Grad-CAM	48.390	10.894	37.496	25.851
	Grad-CAM++	45.044	12.528	32.516	22.331
	XGrad-CAM	49.029	10.761	38.268	26.223
	Eigen-CAM	48.925	10.547	38.378	25.765
	Layer-CAM	48.125	10.444	37.681	26.002
	CAMERAS	44.548	9.091	35.457	26.153
	Relevance-CAM	49.296	10.043	39.253	25.980
	IAA	43.733	10.116	33.617	24.876
	<b>Holistic-CAM</b>	<b>49.569</b>	<b>8.792</b>	<b>40.777</b>	<b>26.345</b>
	Grad-CAM	11.226	14.503	-3.277	-3.696

Shallow Layer [Layer23]	Ins↑	Del↓	Over-all↑	ROAD↑	EPG↑
Deep Layer [Layer43]	Grad-CAM	20.927	10.394	10.533	12.468
	XGrad-CAM	18.142	9.216	8.926	12.036
	Eigen-CAM	42.230	7.276	34.954	25.417
	Layer-CAM	39.160	5.882	33.278	25.656
	CAMERAS	39.994	6.972	33.022	26.008
	Relevance-CAM	32.180	8.487	23.693	19.995
	IAA	14.301	7.552	6.749	16.236
	<b>Holistic-CAM</b>	<b>45.001</b>	<b>6.969</b>	<b>38.032</b>	<b>26.252</b>
	Grad-CAM	11.226	14.503	-3.277	-3.696
	Grad-CAM++	20.927	10.394	10.533	12.468