



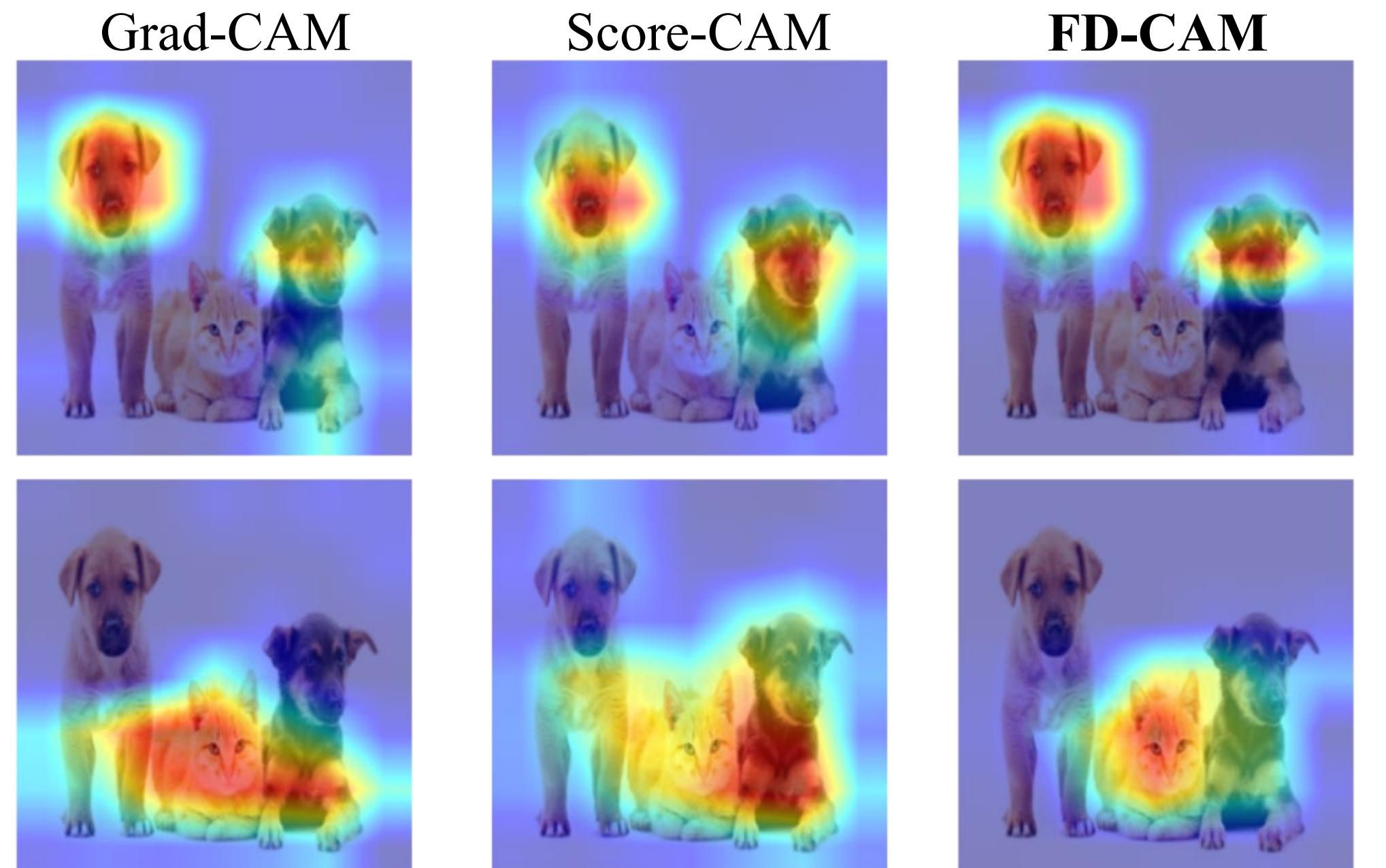
FD-CAM: Improving Faithfulness and Discriminability of Visual Explanation for CNNs

Hui Li, Zihao Li, Rui Ma, Tieru Wu
Jilin University, China

CVPR 2022
XAI4CV Workshop

INTRODUCTION

- Class activation map (CAM) has been widely studied for visual explanation of convolutional neural networks.
- Existing gradient and score based weighting schemes show superiority in ensuring either the discriminability or faithfulness, but they cannot excel in both properties.



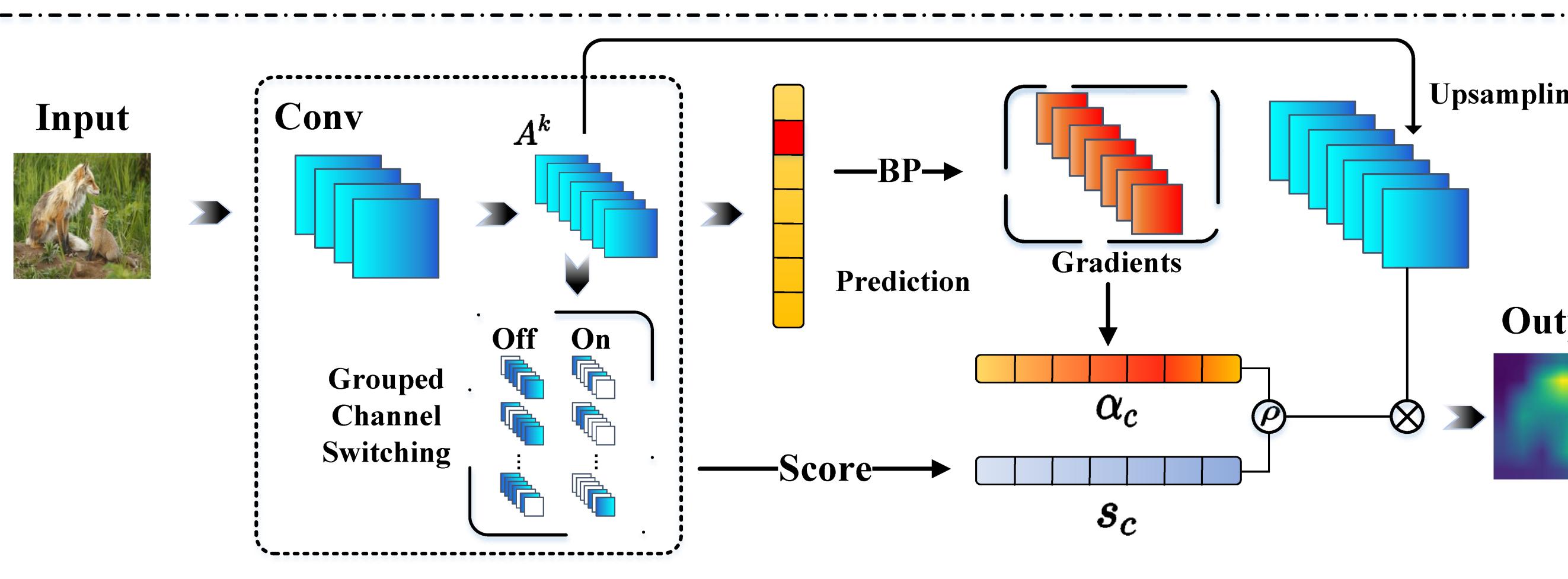
CONTRIBUTIONS

- We propose FD-CAM, a novel CAM weighting scheme which **combines the gradient and score based weights** to improve the faithfulness and discriminability of visual explanation for CNNs.
- We introduce the **grouped channel switching** which perturbs groups of channels simultaneously.
- The new scheme for CAM weighting is proposed and the code is available on GitHub
<https://github.com/crishhh1998/FD-CAM>.

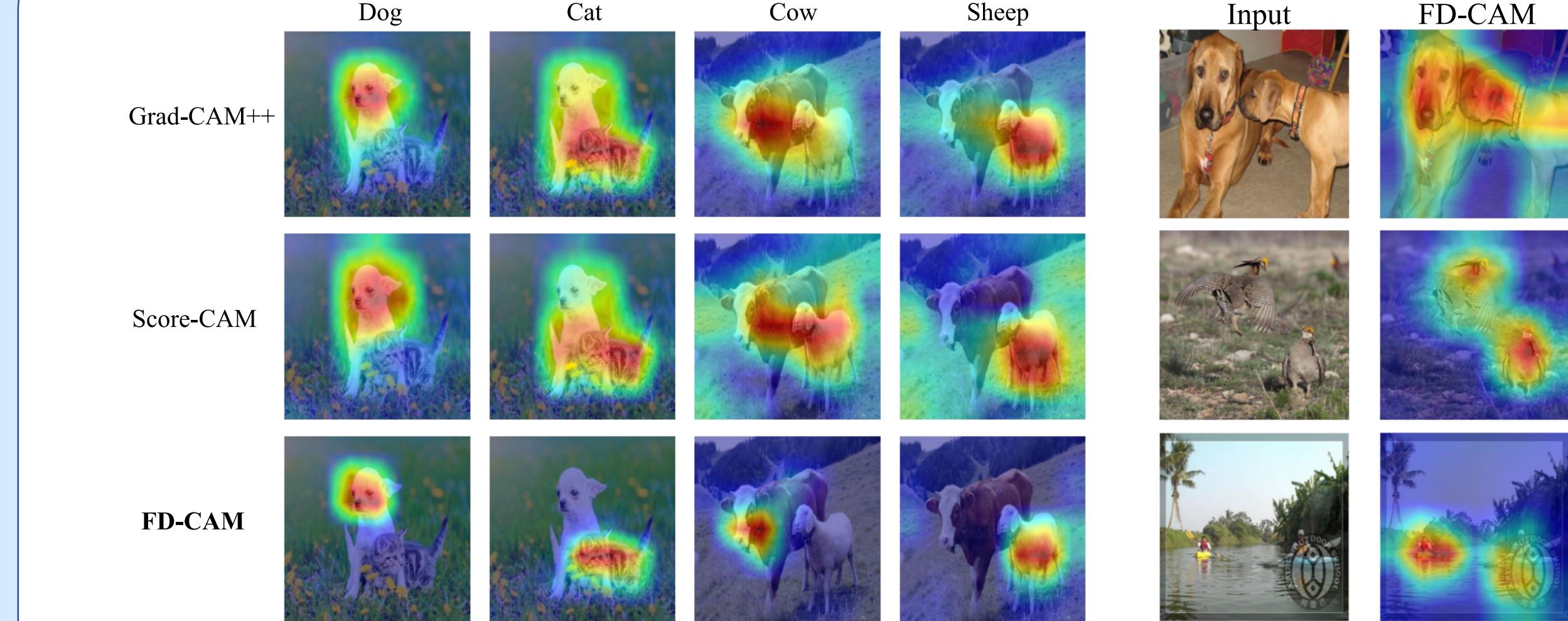
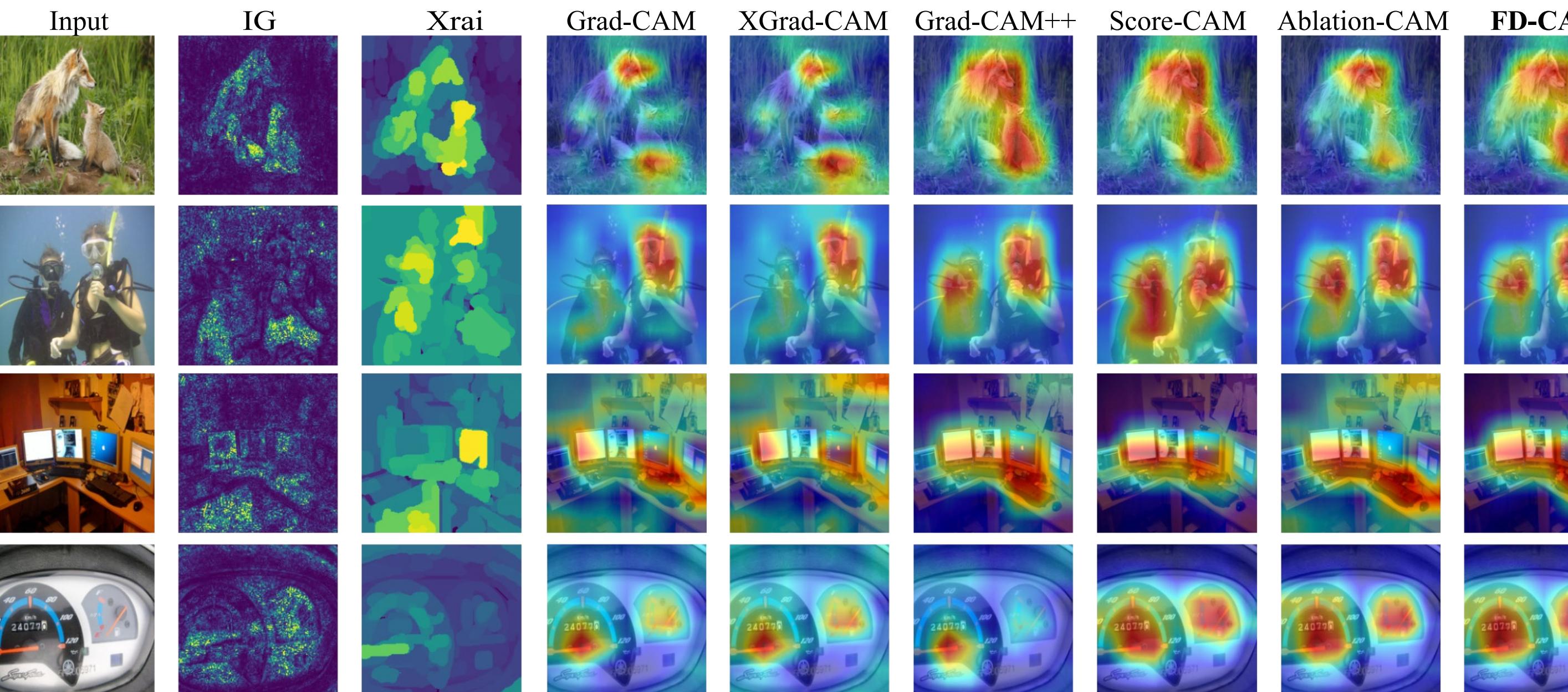
METHOD

- FD-CAM scheme combines the gradient-based with score-based weights.
- Score-based weights: perform grouped channel switching operation to find the importance of each channel. The reason is, in a particular convolutional layer, the features in different channels have high similarities.
- Combination of gradient and score based weights to define final weights:

$$\omega_c^k = \rho(\alpha_c^k, s_c^k) = \hat{\alpha}_c^k e^{\hat{s}_c^k} - b$$



RESULTS



Methods	Insertion ↑	Deletion ↓	Overall ↑	Acc(%) ↑
Grad-CAM [13]	0.5357	0.1117	0.4240	81.20
Grad-CAM++ [17]	0.5321	0.1088	0.4233	81.91
XGrad-CAM [28]	0.5464	0.1072	0.4392	80.72
Score-CAM [14]	0.5422	0.1059	0.4363	78.46
Ablation-CAM [23]	0.5502	0.1013	0.4489	58.19
Layer-CAM [29]	0.5389	0.1021	0.4368	81.77
Group-CAM [22]	0.5397	0.0921	0.4476	-
FD-CAM	0.5534	0.1001	0.4533	83.70

CONCLUSIONS

- FD-CAM shares the advantages of both gradient and score based CAM in producing faithful and discriminative visual explanation of CNNs.
- **Faithfulness:** FD-CAM's results can accurately highlight the regions related to the model's decision to a greater extent.
- **Discriminability:** FD-CAM outperforms in distinguishing different classes with different prediction probabilities.

ACKNOWLEDGEMENTS

This work is supported by the National Key Research and Development Program of China (Grant 2020YFA0714103) and the National Nature Science Foundation of China (Grant 61872162).