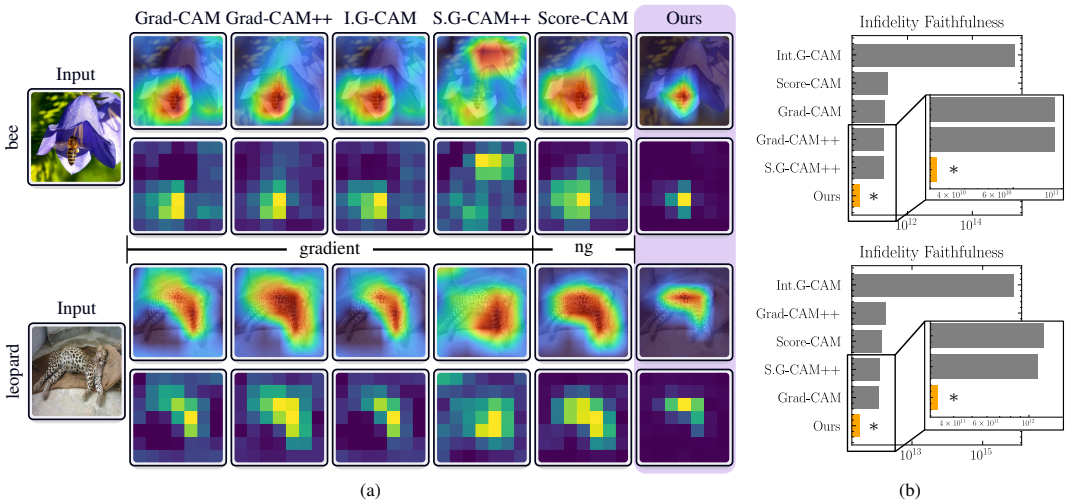
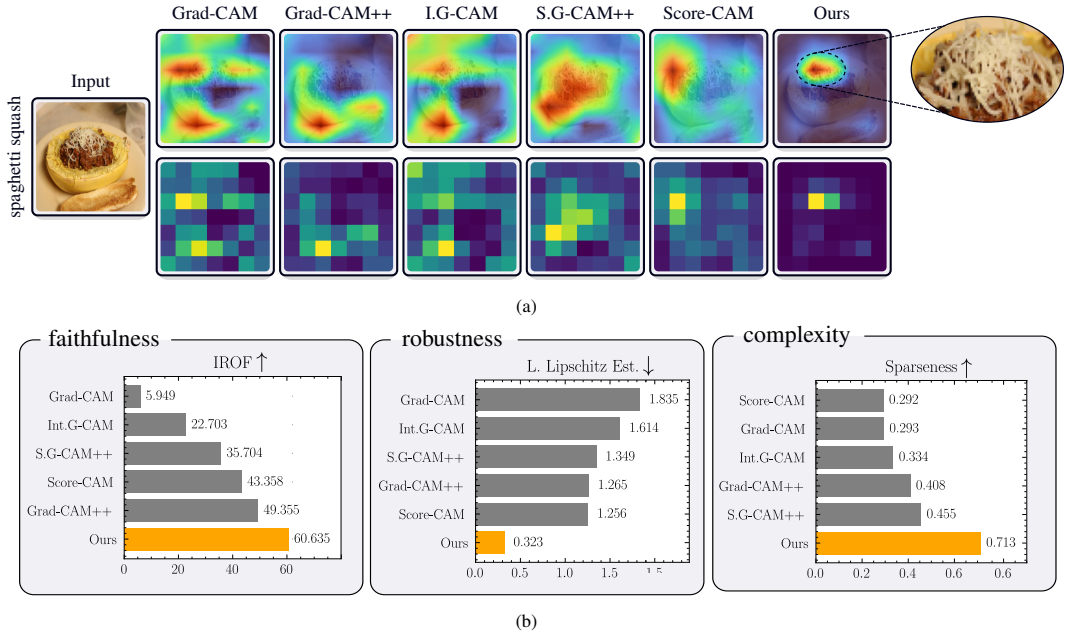


(a) (b)

Figure 1: Comparison of saliencies generated by different gradient- and non-gradient-based methods. Figure 1a shows the superimposed (top row) and raw coarse saliencies (bottom row) generated by each method. Figure 1b presents the Infidelity scores [**?**] (using log-scale) for the different methods. While baseline methods are noisy with low localization, our method produces sharper, more localized explanations, outperforming even non-gradient-based techniques, and resulting in significantly lower infidelity scores (fig. 1b).



(a)

(b)

Figure 2: Comparison of saliencies generated by different gradient- and non-gradient-based methods. Figure 2a shows the superimposed (top row) and raw coarse saliencies (bottom row) generated by each method. Our method consistently produces more focused and sharper saliencies compared to both gradient-based and non-gradient-based methods (e.g., Score-CAM). Figure 2b demonstrates that our approach concurrently improves key xAI properties: (i) faithfulness, (ii) robustness, and (iii) complexity, significantly outperforming even non-gradient-based methods.