

This study addresses generating counterfactual explanations with multimodal information. Our goal is not only to classify a video into a specific category, but also