

Abstract

The aim of the research. In this paper, the approach to search for multiple explanations of the CNN image classification case is proposed. **Research results.** The core of the method is recursive division (RD), that performs the perturbation of the input image with hiding different rectangular parts. The explanation is represented as a complementary images pair (CIP): two images that allow us to visualize the parts of the image which are important enough to change the class of the input image when hidden and at the same time are important enough to preserve the initial classification result when visible. The parameters of RD method are discussed to choose the criteria to stop the processing when few explanations are found or the further processing requires too much time and/or memory resources. Two approaches to merge multiple CIP back to single explanation using SLIC segmentation were proposed. They allowed us to reduce the useful image explanation area and sometimes find more visually attractive CIP compared to previous RD implementations. Such merging is not strictly required just multiple CIP explanations are good enough for analysis of the CNN. **Conclusion.** The implementation of the proposed approach for cats and dogs breed classification problem was compared with other popular methods like RISE and Grad-CAM, the benefits and drawbacks are discussed. The performance analysis confirmed the advantage of the proposed methods as they are comparable or faster with known and allow us to find multiple explanation images.

Keywords:

Image Classification, CNN, AI, Black-box, Explainability, Image Perturbation, Hiding Parts, Recursive Division (RD), Complementary Image Pair (CIP)