Adversarial Perturbation on MRI Modalities in Brain Tumor Segmentation

summarize this text:

In recent years, deep learning networks and especially convolutional neural networks (CNNs) have achieved remarkable success in many computer vision areas, including object recognition [1], [2], semantic segmentation [3], [4] [5], depth estimation [6], object detection [7], [8], etc. Instead of crafting features by humans, CNNs use convolutional layers to automatically extract the features from the input images and provide end-to-end solutions to the perceptional task. Because CNN-based solutions have more accurate feature extraction and are more convenient in the processing pipeline, convolutional neural networks have also been widely used for biomedical segmentation tasks including lung segmentation [9], [10], brain tumor segmentation [11], etc. The goal of brain tumor segmentation is to detect and localize tumor regions by comparing the tested brain tissue images to the normal brain tissue images [12]. The ground truth brain tissue images are labeled by a medical professional on special medical images such as X-ray or MRI images. Automatic brain tumor segmentation has been used to help medical personnel, reducing the time necessary for identification of abnormal regions. Automatic segmentation methods will be critical for early tumor prescreening, especially when doctors need to examine a large number of biomedical images. Currently, magnetic resonance imaging (MRI) is a major type of biomedical imagery in brain tumor analysis, monitoring and surgery planning. Compared to the methods using handcrafted features, methods based on CNNs use a set of convolutional filters that can extract the convolutional features directly from the input data, providing end-to-end solutions to MRI images. Many state-of-the-art automatic brain tumor segmentation methods have been developed in the recent years, such as U-net [13] and V-net [12], [14]. Although both X-rays and CT images can be used for biomedical research, MRI images are the major data source type for these neural networks. Research on adversarial attacks has shown that typical convolution neural networks have a universal vulnerability to these attacks. Similar to other CNN-based applications, semantic segmentation applications have also been proven to be vulnerable to adversarial sample attack [15], [16]. However, compared to other semantic segmentation applications, brain tumor segmentation has several unique characteristics. Because brain tumors have different sizes, shapes, and locations in different patients, doctors and other medical personnel always use several modalities of MRI images to help tumor region segmentation and labeling. Different modalities have differences in the pixel intensity and contain different information. The comprehensive consideration of multiple modalities provides tumor tissues at multiple intensity levels for analysis by doctors. There are four types of modalities of MRI images: T1 (spin-lattice relaxation), contrast-enhanced T1 (T1ce), T2 (spin-spin relaxation) and FlAIR (fluid-attenuated inversion recovery). Each modality corresponds to grayscale images that highlight different kinds of tissue. In current brain tumor segmentation methods, all four modalities are used jointly for computation during the model training process. As a result, it is beneficial to investigate how adversarial perturbation affects the brain tumor segmentation methods and to elucidate the impact of adversarial perturbation in different modalities. The research described in this paper can be useful for the mitigation of many threats:

1. As the images are acquired from MRI equipment, system failures and human error may cause imperfect MRI images, possibly leading to the errors in the segmentation results.
2. As MRI images are valuable personal information, they are vulnerable to cyber attack and can be deliberately altered by adversaries. In this paper, we first evaluate the adversarial perturbation effect on the current automatic brain tumor segmentation methods in terms of segmentation accuracy and then investigate the performance reduction for the adversarial attacks on each modality. This paper also presents an explanation for the accuracy differences among the four modalities. The adversarial training recommendation will be given in the end of the paper. The rest of the paper is organized as follows: Section II introduces the background of U-net-like neural networks , and the background of adversarial sample attack . Section III introduces the dataset used for this research. Section IV describes the method for the generation of the perturbations for the brain tumor MRI images. Section V shows a typical U-net-like segmentation target model that serves as the attack target. The experimental configuration and results are discussed in Section VI. The conclusion and future work are given at the end of the paper.

Result

The article discusses the use of convolutional neural networks (CNNs) for biomedical segmentation tasks, specifically brain tumor segmentation using MRI images. The article highlights the vulnerability of CNN-based models to adversarial attacks and investigates the effect of such attacks on brain tumor segmentation methods. The article uses a U-net-like segmentation target model as the attack target and evaluates the impact of adversarial perturbations on each modality of MRI images. The article proposes recommendations for adversarial training to mitigate the impact of adversarial attacks. Overall, the article highlights the potential of CNN-based methods for biomedical segmentation tasks while also emphasizing the need to account for the vulnerability of these models to adversarial attacks.

continue

The article emphasizes the importance of automatic brain tumor segmentation for helping medical personnel reduce the time necessary for identification of abnormal regions, especially when doctors need to examine a large number of biomedical images. The article also outlines the unique characteristics of brain tumor segmentation, including the use of multiple modalities of MRI images to aid in tumor region segmentation and labeling. The article proposes the use of U-net and V-net as state-of-the-art automatic brain tumor segmentation methods based on CNNs. The article also highlights the potential risks associated with imperfect MRI images, system failures, human errors, and cyber-attacks, and proposes recommendations for adversarial training to mitigate these risks. Finally, the article suggests future research in this area should focus on investigating the impact of adversarial perturbations in other biomedical segmentation tasks and developing more robust models to improve the accuracy and efficiency of biomedical image analysis in healthcare.

continue

Overall, the findings of this article highlight the potential of CNN-based methods for automatic biomedical segmentation tasks, particularly in the field of brain tumor segmentation. The article underscores the importance of considering the vulnerability of these models to adversarial attacks and proposes recommendations for adversarial training to mitigate this risk. The article contributes to the broader field of biomedical image analysis by identifying gaps in current research and proposing areas for future research. By addressing these gaps, researchers can improve the accuracy and efficiency of medical diagnostics and treatments, thereby improving patient outcomes and reducing healthcare costs. The article's practical implications for the medical field are significant, and continued research in this area is essential for advancing the field of biomedical image analysis and improving the quality of patient care.