Qeios

Peer Review

Review of: "Omni-IML: Towards Unified Image Manipulation Localization"

Cong Lin[1]

1. Guangdong University Of Finances and Economics, Guangzhou, China

Paper Summary :

This paper addresses the problem of Image Manipulation Localization (IML) with diverse scenarios and proposes a novel unified modeling approach, Omni-IML. Omni-IML can handle manipulation localization tasks for natural images, document images, and face images without requiring specific optimizations for each type. The authors propose (1) a new Modal Gate Encoder, which can automatically select the optimal encoding modality (frequency + visual or pure visual) for input images, (2) an Anomaly Enhancement (AE) technology, which utilizes box supervision to enhance the contrast of manipulated region features while reducing feature noise from joint training; and (3) a new Dynamic Weight Decoder (DWD), which can dynamically select the optimal filters of the decoder, thereby effectively addressing the diversity of manipulation features across different image types. Extensive experiments validate the model's performance on tasks in three different domains (natural images, document images, and face images), showing that the Omni-IML model can achieve or surpass current state-of-the-art models without task-specific fine-tuning.

Paper Strengths

(1) This paper presents a generalist Image Manipulation Localization model capable of unifying multi-task performance, establishing a new paradigm of multi-task unified modeling in this research direction. The motivation behind this paper aligns with the current trend of AGI development, providing a feasible implementation strategy for the field of IML in the AGI era.

(2) The authors introduce a box supervision design to enhance features of tampered regions while suppressing noise during joint training, improving the robustness of the model across different IML tasks.

 doi.org/10.32388/3F8SU6

(3) In multi-task joint training scenarios, features from disparate image types can introduce noise. The proposed AE module mitigates this issue by amplifying the feature contrast within manipulated regions, resulting in improved multi-task adaptability.

(4) The proposed DWD module can perform adaptive selection of sample-specific decoder filters that helps the model handle diverse tampering features and reduces confusion in the unified training process.

Major Weaknesses

(1) The use of frequency domain features in the Modal Gate Encoder enhances the model's adaptability. However, a crucial consideration is that when the frequency domain information of the input image is overly complex or corrupted by noise, it can potentially degrade the model's performance. Please clarify this.

(2) The model's performance in the paper is validated on specific high-quality datasets (such as tampCOCO, SACP, and FaceShifter). It is unclear whether these datasets can cover the complex scenarios of the real world. Can the method in this paper be applied to a broader range of image manipulation datasets, such as IMD2020, DSO-1, and Korus? Furthermore, is the method feasible for manipulated images generated by generative methods, such as the AutoSplicing and OpenForensics datasets?

Minor Weaknesses

- Several font sizes in Figure 2 are too large; it is recommended to reduce them.

Declarations

Potential competing interests: No potential competing interests to declare.