TECH OFFER

Improving Explainable Artificial Intelligence For Degraded Images

KEY INFORMATION

TECHNOLOGY CATEGORY:
Infocomm - Video/Image Analysis & Computer Vision
Infocomm - Artificial Intelligence

TECHNOLOGY READINESS LEVEL (TRL): TRL5
COUNTRY: SINGAPORE
ID NUMBER: TO174361

OVERVIEW

One use of AI, including deep learning, is in prediction tasks, such as image scene understanding and medical image diagnosis. As deep learning models are complex, heatmaps are often used to help explain the AI's prediction by highlighting pixels that were salient to the prediction.

While existing heatmaps are effective on clean images, real-world images are frequently degraded or 'biased'-such as camera blur or colour distortion under low light. Images may also be deliberately blurred for privacy reasons. As the level of image clarity decreases, the performance of the heatmaps decreases. These heatmap explanations of degraded images therefore deviate from both reality and user expectations.

This novel technology-Debiased-CAM-describes a method of training a convolutional neural network (CNN) to produce accurate and relatable heatmaps for degraded images. By pinpointing relevant targets on the images that align with user expectations, Debiased-CAMs increase transparency and user trust in the AI's predictions.

TECHNOLOGY FEATURES & SPECIFICATIONS

- Debiased-CAMs are effective in helping users identify relevant targets even on images affected by different clarity levels and multiple issues such as camera blur, poor lighting conditions and colour distortion. The AI's prediction also becomes more accurate.
- As the model is trained using self-supervised learning, no additional data is needed to train it.
- The training for Debiased-CAM is generalisable, and thus applicable to other types of degraded or corrupted data and other prediction tasks such as image captioning and human activity recognition.

POTENTIAL APPLICATIONS

Used to train a convolutional neural network (CNN) to produce accurate and relatable heatmaps for degraded images. By pinpointing relevant targets on the images that align with user expectations, Debiased-CAMs increase transparency and user trust in the AI's predictions. It also increases the ability of meeting regulatory standards to deploy CNN models in the following applications, where explainable AI is required.

- Healthcare, eg. Radiology
- Autonomous Vehicles

UNIQUE VALUE PROPOSITION

- Produces accurate, robust and interpretable heatmaps for degraded images
- Works on images with multiple degradation levels and types such as blurring and improper white balance
- Agnostic to degradation level, so that enhancement can be applied even when the level is unknown
- Perceived by users to be more truthful and helpful as compared to current heatmaps distorted due to image degradation
- Method of training can be applied to other degradation types and prediction tasks