CROCODILE: Crop-based Contrastive Discriminative Learning for Enhancing Explainability of End-to-End Driving Models

Chenkai ZHANG, Daisuke DEGUCHI, Jialei CHEN, Zhenzhen QUAN, and Hiroshi MURASE

Nagoya University, Nagoya Aichi, Japan, zhang1354558057@gmail.com

1 Explanations showed in Experimental Results and Discussion

In the supplementary materials, we show the explanations discussed in the paper. We show 4 groups of explanations:

- For **Table. 1** in **Section 5.1**, to demonstrate the effectiveness of CROCODILE across different backbones, we show the explanations in Fig. 1.
- For **Table. 3** in **Section 5.2**, to demonstrate the effectiveness of CROCODILE across different E2EDMs, we show the explanations in Fig. 2 and Fig. 3.
- For **Table. 4** in **Section 5.3**, to analyze the effectiveness of various versions of the CROCODILE, we show the explanations in Fig. 4.
- In addition, we demonstrate how the explanations change during the training of CROCODILE. Each backbone is trained by CROCODILE for 200 epochs, saved every 25 epochs, resulting in eight backbones with different levels of training. Then, we fine-tune the E2EDMs based on this series, resulting in 8 E2EDMs. Since we found that CROCODILE performs best with ResNet18 and CCnet, we show the explanations of 8 Ours-R18-CCnets in Fig. 5.

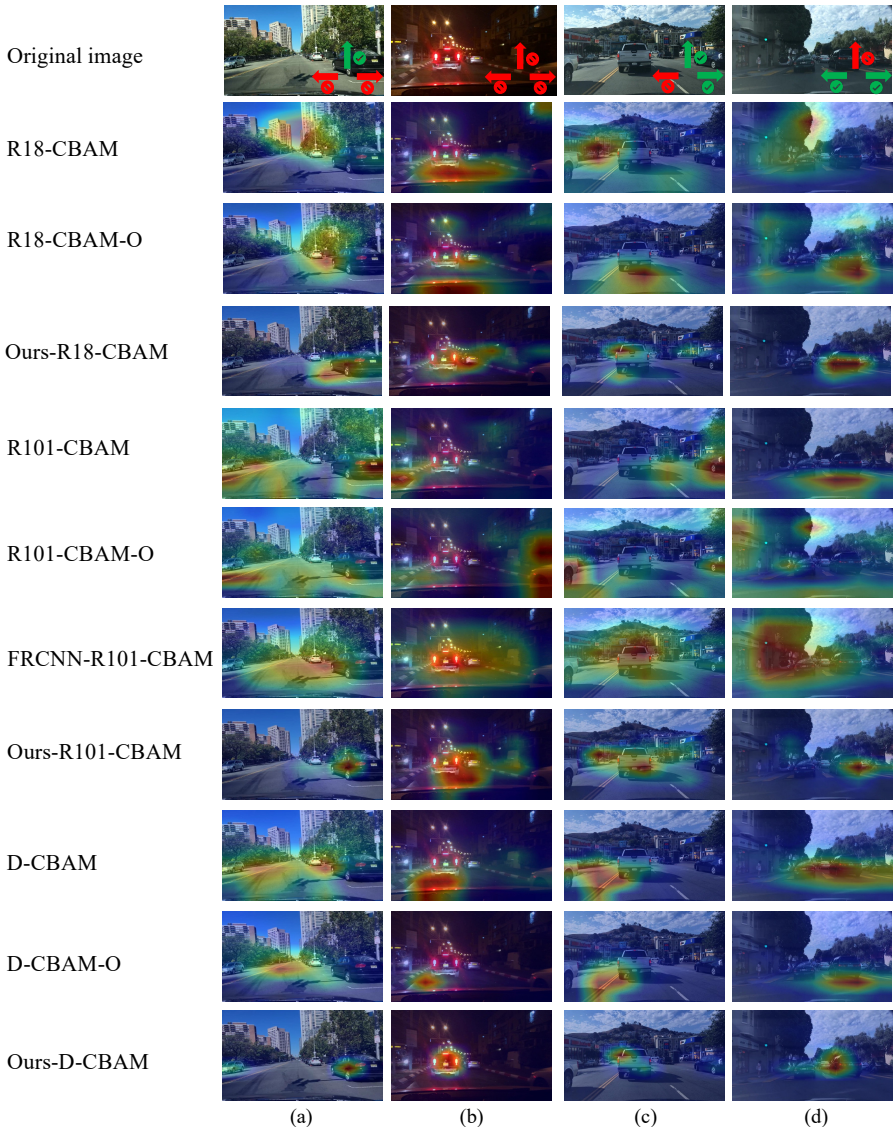

Fig. 1: To demonstrate the effectiveness of CROCODILE across different backbones, we show the explanations discussed in **Table. 1**. For each backbone, the E2EDMs from the CROCODILE have a stronger ability to utilize important object features compared to the baselines. FRCNN-R101-CBAM tends to utilize more object features instead of focusing on important object features.

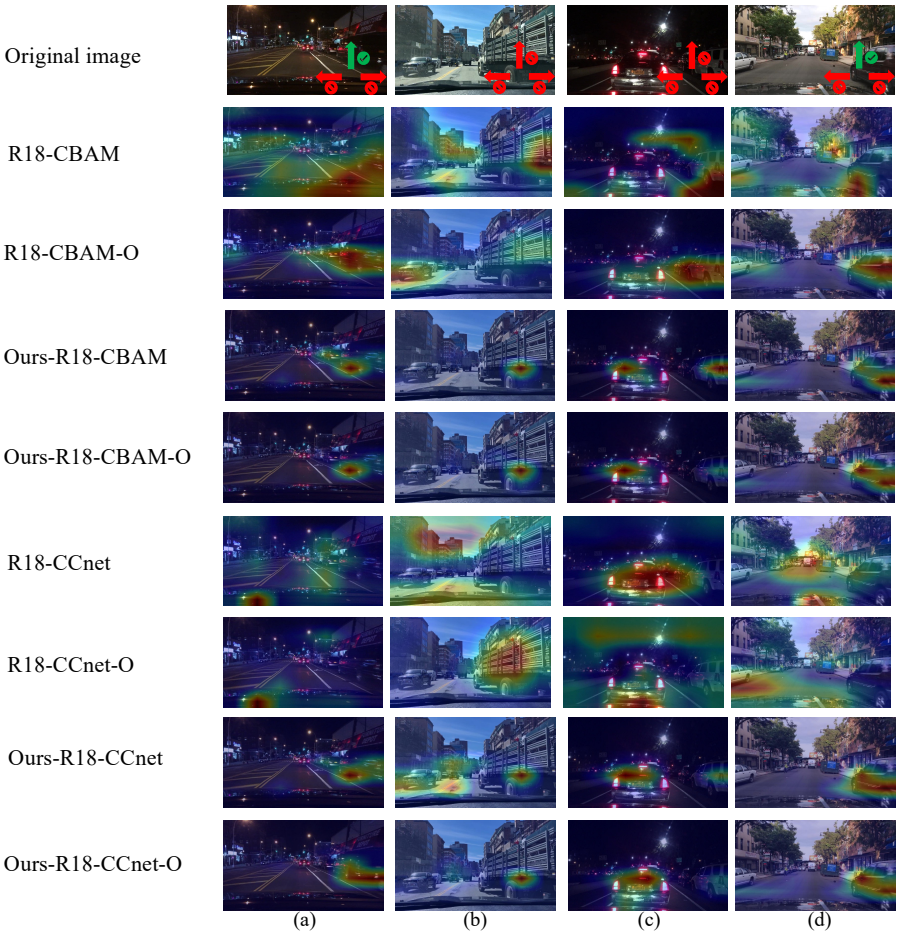

Fig. 2: To demonstrate the effectiveness of CROCODILE across different E2EDMs, we show the explanations discussed in **Table. 3**. For each E2EDM, the E2EDMs from the CROCODILE have a stronger ability to utilize important object features compared to the baselines. E.g., R18-CBAM-O focuses more on objects compared to R18-CBAM, however, Ours-R18-CBAM, and Ours-R18-CBAM-O are better at focusing on important object features.

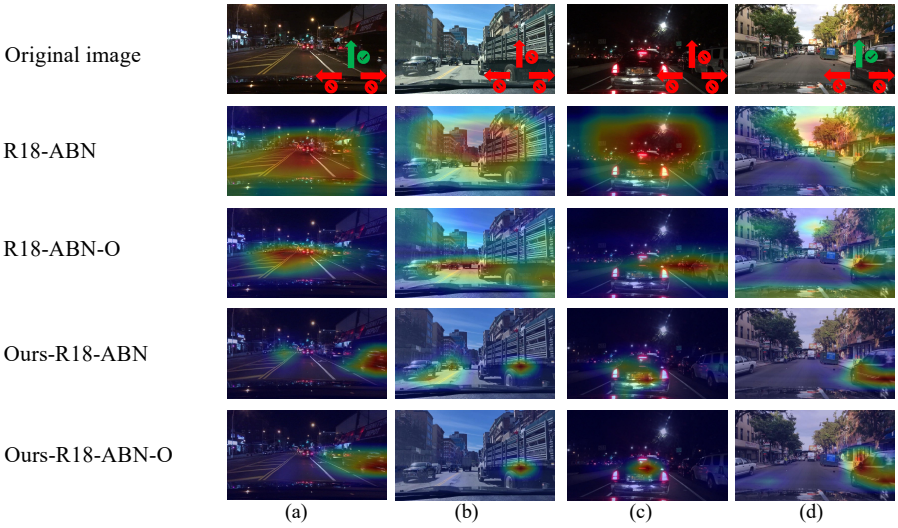

Fig. 3: This is a continuation of the Fig. 2.

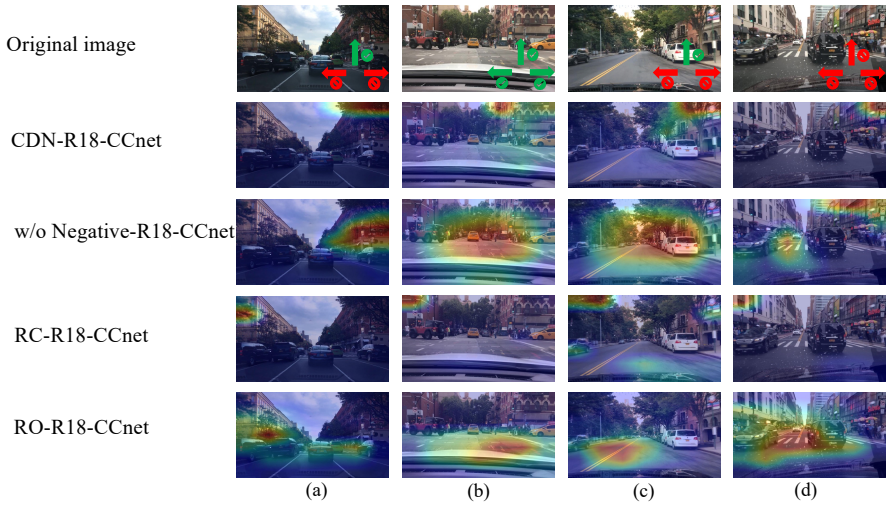

Fig. 4: To analyze the effectiveness of the CROCODILE, we remove or replace certain components of the CROCODILE and observe whether the modified CROCODILE remains effective, we show the explanations discussed in **Table. 4**. The w/o Negative-R18-CCnet and RO-R18-CCnet still have the ability to extract object features, but their ability to focus on important objects is weakened. The CDN-R18-CCnet and RC-R18-CCnet completely fail to understand the driving scene.

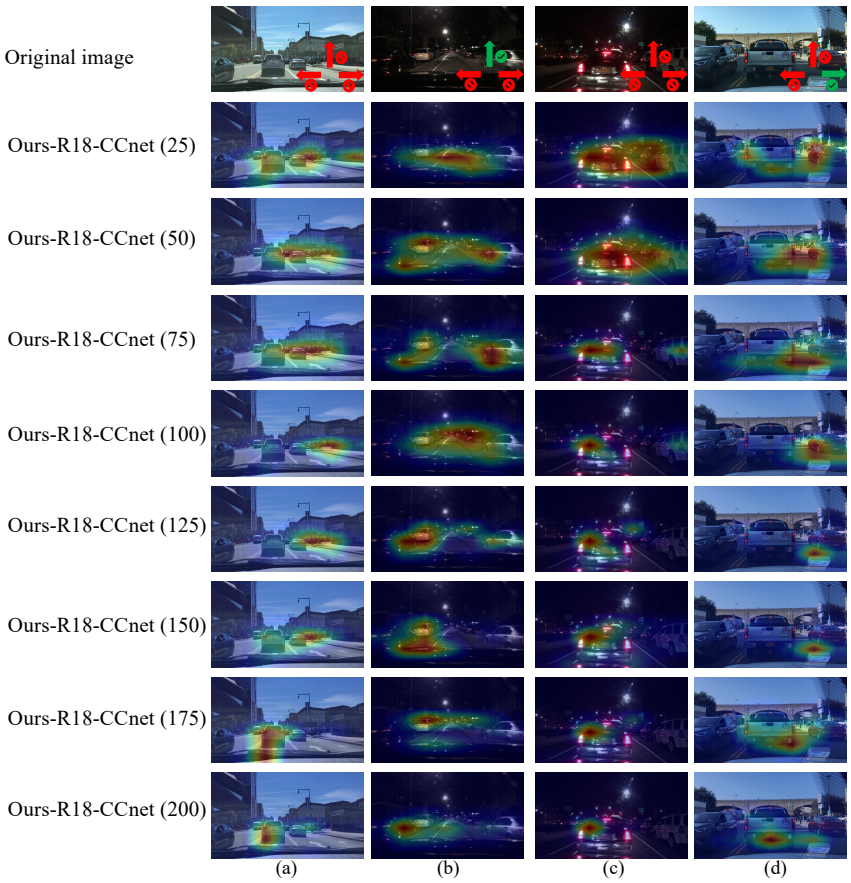

Fig. 5: To analyze the effectiveness of the CROCODILE, we demonstrate how the explanations change during the training of CROCODILE. As the training duration increases, the highlighted area gradually shrinks to the important objects, *i.e.*, the E2EDMs gradually utilize concise important object features to make predictions.