

We would like to thank the reviewer and the associate editor for their time and constructive comments, which enable us to greatly improve the quality of our manuscript. We also provide a marked-up copy of our manuscript that highlights changes made to the original version. Next we provide point-by-point responses to the review comments.

(1) *In Page 6, L229, the authors mention that “...c is the value that the generator can fool the discriminator”, but c does not appear in Eq. (4). Should c be replaced by a? A similar problem occurs in Eq. (8).*

Reply:

Thank you for pointing out this issue. We have fixed it in our new version.

(2) *In Figure 2, the input arrow in front of the Encoder-fg and Encoder-bg is not connected to the encoder's first layer, which is ambiguous.*

Reply:

Thanks. We have modified Figure 2 to make sure it is clear.

(3) *In Page 7, L251-253, the authors mention that “...to extract the desired features better, the inputs of two encoder paths are both determined to be 1: 2.”, the authors are suggested to explain this in detail.*

Reply:

Thank the reviewer for this valuable suggestion. We have explained this problem in detail in our revision. The feature characterization between foreground and background is distinct, therefore, two encoding paths are employed to extract the feature maps of them, respectively. Moreover, in the ideal foreground, the thermal radiation target is remarkable, and the texture details of the ideal background are rich. In order to achieve the effect as far as possible, two times the infrared image foregrounds and one time the visible image foreground are input to the foreground encoding path, while the background encoding path takes one time the infrared image background and two times the visible image backgrounds as input. Finally, the feature maps of the foreground and background are concated as the decoder's input to generate a fused image.