Improving Accuracy of Respiratory Rate Estimation by Restoring High Resolution Features with Transformers and Recursive Convolutional Models

Author Responses to Reviewers' Comments

We would like to thank the Reviewers for their time and effort required to review the submitted manuscript. We sincerely appreciate all the valuable comments and suggestions, which helped us to improve the quality of the manuscript. Below we summarize all introduced changes.

1. Paper Organization and Sections Naming

"The paper organization and title of some sections need to be improved. The authors adopt some existing deep network models. It is difficult to differentiate the contribution of proposed method and the existing model."

The names of related work, problem statement and methodology sections have been modified to better capture the main outcome of each of them. In addition, the summary of contributions in the introduction section has been also revisited to make the proposed techniques more clear.

2. Limitations of the Proposed Method

"In section 3.2, the sensor is set up at a distance of 1.2m from volunteer face. This is not a natural setting. At this close range, I think it is a strict restriction on the user." "Experiments were done in a limited/controlled environment. However in real applications, many factors could influence the estimation, such as environmental temperature, head angle or face angle to camera, breathing rate range (...) These limitations should be clearly highlighted to avoid delivering overoptimistic impression and overclaiming."

We agree that it's very important to specify these limitations. An appropriate paragraph has been added to the discussion section "Although the results are very promising, they are preliminary and should be further verified in the future work. First of all, it's very important to perform similar analysis in less controlled environments, as various factors can influence the reliability of the estimation, i.e. camera angle, body position, environment conditions, etc. Secondly, the presented study addresses only single person setting, at a close proximity to a sensor, due to target applications, such as vital signs deployed at the border control, computer stations, etc. However, real-life scenarios would require less strict restrictions on the user, what should be further analysed. "

3. Reference RR Estimation Methods

"The experiment is not sufficient. There is no comparison with existing respiration monitoring system, please refer to [1][2]."

Thank you very much for your suggestion. It's important to note that this study doesn't aim at proposing better RR evaluation techniques. Instead, the main goal of our research was to evaluate the possibility of improving accuracy of vital signs extracted in a non-contact way by enhancing the texture and details of low resolution thermal sequences. Thus, we don't compare different respiration monitoring systems, but analyze how various resolution enhancement techniques affect the accuracy of the exemplary RR evaluation method, previously verified in the literature to produce satisfactory estimation results. We will continue a further analysis of the influence of resolution enhancement on accuracy of other RR methods in future studies. We've also added some ideas for future work, indicating the need for verifying the method against ground truth measurements obtained with professional devices.

4. TTSR Pros and Cons

"If the reviewer understands correctly, TTSR requires the use of reference images, which would be a clear practical limitation when there is no reference high-resolution images for training (...) The authors also mentioned a bit why they prefer TTSR, but this was not very well discussed, without clear support of evidence. TTSR makes use of the reference images, so please discuss whether the comparison is fair even."

We appreciate this valuable feedback and completely agree that requirement of the reference image is the limitation of the reference-based SR method. That's why we performed additional experiments with images from different domains (visible light) which were representing different objects used for both model training and during inference for transferring textures. As presented in Table 1, such an approach led to the second best RMSE result, what might allow for eliminating the need for acquiring HR data and using other images as a reference instead (e.g. from existing datasets, such as ImageNet). We've added this description to the discussion section. In addition, in the discussion section we also indicate the need for fine-tuning the RefSR model on the thermal dataset in future studies in order to provide more in-depth and fair comparisons of different NN architectures.

5. Analysis of PSNR

"PSNR is hard to interpret, please describe more details."

Thank you very much for this suggestion. We've added results and interpretation of the SSIM metric which correspond to the perceived quality of the image and thus should be more intuitive and easier to interpret.