Rebuttal

We sincerely thank the reviewers for time and insightful feedbacks! We address the issues in detail as follow:

[Review1-Q1]: The experimental results section could do with more details on the number of hand-drawn sketches.

[A1.1]: We invited \*\*one\*\* expert with well-trained drawing skills to draw \*\*12\*\* portrait sketches for testing. Also, we invited \*\*20\*\* graduate students without drawing skills to \*\*200\*\* draw freehand sketches. These details will be added in revised paper.

[Review1-Q2]: Are results in Figure 6/7 representative of the entire set, or only the best results?

[A1.2]: Basically, most of our results on both types of hand-drawn sketches surpass those of both baseline methods in generating fine textures and realistic face shapes. We selected results shown in Figure 6/7 to better demonstrate our model’s effectiveness on spatially adjusting balance between realism and conformance, which is one of the key ideas of the proposed method. See [A3.3] for further explanation on this idea. We will emphasize this idea in the descriptions of Figure 6/7.

[Review2-Q1]: Comments on typos and minor errors.

[A2.1]: Typos and errors are to be modified in revised paper.

[Review3-Q1]: Can the proposed solution generalized to other types like animals, cars, etc. how and why?

[A3.1]: Face images in the used dataset are aligned, and the structure of faces are well-defined. These two characters reduce the difficulty of our task (though it is still very challenging). The proposed solution might be easy to be generalized to datasets with aligned images of other types, such as animal face dataset [MUNIT, Huang et al. ECCV 2018], shoes dataset and handbag dataset [pix2pix, Phillip et al. CVPR 2017]. However there potentially exists some challenges in generalizing to unaligned dataset or dataset with multiple object classes, like ImageNet, due to the diversity of structures, positions and scales.

[Review3-Q2]: Comparison with paper Chen et al.

[A3.2]: It is not able to compare with this paper, since this paper, accepted by Siggraph 2020, is not published and, as far as I know, is not available at arXiv until June 2020, AFTER the submission date of our paper.

[Review3-Q3]: 'We argue that the balance between the realism and the conformance differs from one position to another across the face image'. This is not clear. Can you explain why?

[A3.3]: We argue that the quality of each part in one common hand-drawn sketch differs from each other. The balance moves forwards to the conformance at a well-drawn part with realistic shape to meet the desire/tendency of user, while moving forwards to the realism at a poor-drawn part to ensure the quality of the generated image.

[Review3-Q4]: Eq.8 is not clear.

[A3.4]: Eq.8 is to be modified in revised paper.

[Review3-Q5]: The proposed solution is very expensive in term of complexity because it requires treating each facial feature locally first and then the face as a whole? can you elaborate more on this?

[A3.5]: The complexity is not as large as it appears since the number of channels of feature maps in SAP is small (48).

[Review3-Q6]: More datasets and comparison are needed for the evaluation.

[A3.6]: More experiments are to be added in revised paper.

以上字符数：3212 （Rebuttal要求最大字符数是5000）

草图的具体数量还要根据最后的收集的数量修改。

Discuss:

1. 这句话应该用什么时态？

* Xxxx is to be modified in revised paper.
* Xxxx is modified…
* Xxxx has been modified…
* Xxxx will be modified…

1. 问题[Review3-Q6]要不要回答？回答的话，里面要不要指明将加入什么实验？