Dear Editor,

Find below our point-by-point response to the final minor comments.

Yours sincerely,

Arjen Alink

Reviewer Comments:

Reviewer 2
The authors have addressed my comments on the original version of the manuscript. I think that the revisions have improved the paper, and that the findings will be of interest to researchers studying vision and/or autism. I have only a couple of comments related to a revision described here:

[taken from the authors' response to the original reviews...]
Action: We now more explicitly point this out in the results section by including the following text in the 4 th paragraph of the results section:
“Visual inspection of images reconstructed from the 200 features with the highest FDi values revealed that these images were easier to recognize than images reconstructed from the 200 features with the lowest FDi values (shown for one exemplary image in Figure 1c and for all images in the supplementary figure). This observation, however, does not quantitatively validate the efficacy of our method. Therefore, we have, in addition, validated our method by assessing the replicability of the pattern of FDi values (across all 10,000 features) between participants. This analysis revealed significant replicability of FDi patterns across participants (Pearson rho = .081, p < .0001, permutation-based test, Figure 1c).”

[my comments...]
First, I believe that the authors are referring to Pearson r, not Spearman's rho. Correct? If so, they should change rho to r here (and a few other places in the manuscript).

>> We have now changed rho to r throughout the manuscript and figures

Second, the correlation of 0.081 suggests that the reliability of FDi values across participants was low. Certainly, a correlation like this generally would not be used to say a measure was "reliable", despite the fact that it was statistically significant. Now I am not sure that I would expect this correlation to be very high. My understanding is that there were many features that could be deemed as having a "high" FDi. I suspect that each subject uses only a subset of these features to perform the task, and these subsets could differ significantly across observers. This between-subject variation would, I think, reduce the correlation measured by the authors. But perhaps a different kind of correlation would be more meaningful. For example, would it be possible to classify each feature as "high" FDi (e.g., the top 20% coded with a 1), "low" FDi (e.g., the bottom 20% coded with -1), or moderate FDi (e.g., the remaining 60% coded with 0) and then calculate the correlations of those classifications across observers? Would that kind of correlation would be a better index of consistency across observers? In any case, I don't think that the authors should present this low correlation as evidence of "reliability". It is significantly different from zero, yes, but the low correlation suggests that FDi did vary a lot among observers.
>> We have toned down this statement by removing the word ‘reliably’ on page 6 and 9

Despite my comments about the low value of r, my impression is that this is an original and intriguing paper. Certainly it is a study that would inspire discussion and perhaps follow-up experiments in my research group.

Reviewer 1
I am satisfied that the authors have addressed my comments sufficiently. This is a really interesting study and addition to the literature.