Review Access-2024_51179

Manuscript title: Visibility Aware In-Hand Object Pose Tracking in Videos with Transformers.

Authors: Prof. Phan Xuan Tan et al.

General comments: This is my first review of the paper. iI appears that the paper has been rewied by other experts and the authors have corrected/ improved the contents.

My comments are as follows: -

1. The paper is very lengthy. The authors have unnecessarily referred historical papers with no relevance to current work. Consider omitting it.

2. The idea of introducing 'visibility base function is good' and appears to improve the performance noticeably. I see it as 'best effort approach' with current state of occlusion.

3. On page 28 the authors say that the visibility estimation 'is trained' with binary cross entropy method'. it is unclear (not likely) on how they got the training data. In absence of it, binary- cross entropy is very unreliable.

4. Rational behind Sigmoid activation, on page 28 and multi layered perception (MLP), on page 29 is not explained.

Recommendations: Good effort, clarity on above is lacking. Crisp- to the point write-up would have been better. - Weak accept.