REVIEWER 1

Q1. In general, how confident are you in your overall evaluation of this paper?
==> 1: Somewhat Confident

Q2. Rate the importance of this paper
==> 4: Very important

Q2-B. Comment on importance
==> This paper aims to test existing theories about the role of feedback in object recognition in a fully computational instantiation. This topic is of timely importance, and I believe that the authors have made some progress towards this goal. I very much appreciate the degree to which their computational model takes direct inspiration from and connects to existing psychological and cognitive neuroscience literature.
I am still somewhat skeptical about the generalizability of the main findings to new datasets besides Fashion-MNIST and to more complex computational architectures, though this is clearly an avenue for future work. Most importantly, I was also unable to understand some of their secondary results, which are perhaps only accessible to those with more expertise in machine learning. In general, I found that some of the claims of the paper were supported by convincing evidence, while others were made without providing sufficient detail to be evaluated.
REPLY: { Similar trends were observed when we used MNIST. We will test our networks on more complex datasets such as CIFAR-10 in the future. }

Q3. Rate the technical rigor of this paper. Are the results or claims of the paper supported by convincing evidence?
==> 3: Convincing

Q3-B. Comment on technical rigor
==> The key findings of the paper are clearly supported by their main findings, and the authors certainly do not overclaim from their results.
Ultimately, however, I failed to understand some of the secondary results of the paper, making this a bit hard to evaluate. For example, I did not understand why the uninformative cue would boost performative so much relative to the probe-only condition – and this was not discussed. In my opinion, these results were presented without enough detail to understood what they referred to. For example, there are multiple variables with the same name that have different values in the paragraph that beings with “Trends observed in Figure 3…” that I could not understand what they referred to. I also did not understand the results on tuning-based vs. trained-feedback that were described very briefly.
REPLY: { We shortly mention why the uninformative cue provides performance boosts - “This performance boost could be a result of boosting the overall activity of the hidden units (through bias/gain) that provide reliable differences in activity for the object categories, in the case of the images with feature manipulations.” This can be thought of as fine-tuning the network through bias and gain instead of the connectivity weights. We are sorry about the vagueness of the word ‘trends’, although the context seems clear (about the relationships between capacity limits and performance boosts due to the inclusion of the cue). The main point of the comparison with tuning-based feedback was to show that a better alternative

(trained-feedback) exists and raise general questions about the applicability of tuning-based feedback in other systems like the human visual stream. }

Q4. Rate the clarity of presentation. Are the ideas clearly communicated?
==> 3: Average readability

Q4-B. Comment on clarity of presentation
==> I had a hard time understanding the details of this paper, though the general gist was understandable and digestible to a naïve reader. Part of this difficulty came from the use of jargon that has different meanings across different subfields. For example, “representational capacity” and “object processing stream” are psychological/neural constructs that are approximated in this model, but ultimately this is a computational paper. Nonetheless, I think this general line of work holds quite a bit of promise and interest for a general audience.
However, in my opinion, the biggest problem was the authors tried to fit too much into the present paper. This lead to some very brief descriptions of methods/results that made them quite hard to understand.

Q5. In your opinion, how interdisciplinary is the work in this paper? To what extent does the work integrate neuroscience measurements; sophisticated computational techniques, ideas, and models; and tackle rich cognitive theories and phenomena?
==> 4: High

Q6. Overall evaluation of paper
==> 4: Very good

———————————————————————————————————————————

REVIEWER 2

Q1. In general, how confident are you in your overall evaluation of this paper?
==> 3: Very Confident

Q2. Rate the importance of this paper
==> 2: Marginally important

Q2-B. Comment on importance
==> Another "DNN feedback" paper, with one architecture chosen maybe at random, with a task manipulation.
REPLY: { We are not using deep neural networks in the paper. We do not know of any other work explicitly relating the necessity for cue-driven feedback to a specific capacity limit. }

Q3. Rate the technical rigor of this paper. Are the results or claims of the paper supported by convincing evidence?
==> 3: Convincing

Q3-B. Comment on technical rigor
==> Very technical paper, the choice of a particular network is not justified, but it is

ok as it is a short paper. it is difficult to interpret the results
REPLY: { This is the first step in understanding what capacity limits lead to cue-driven feedback being useful. As a first step, we used a simple neural network with one hidden layer. }

Q4. Rate the clarity of presentation. Are the ideas clearly communicated?
==> 3: Average readability

Q4-B. Comment on clarity of presentation
==> Not very clear. the importance of this work is hidden in the myriads of technical details. the authors should clearly write in conclusion what the results and message if this paper are.
REPLY: { As stated in the conclusions, “We found that the feedback boosts performance only if the category-specific features about the objects cannot be fully represented in the base ANN.” We agree that a broader conclusion such as “Cues aid object recognition via feedback only when there is a representational capacity limit on object processing” would have been better suited. }

Q5. In your opinion, how interdisciplinary is the work in this paper? To what extent does the work integrate neuroscience measurements; sophisticated computational techniques, ideas, and models; and tackle rich cognitive theories and phenomena?
==> 3: Medium

Q6. Overall evaluation of paper
==> 3: Good

———————————————————————————————————————————

REVIEWER 3

Q1. In general, how confident are you in your overall evaluation of this paper?
==> 2: Reasonably Confident

Q2. Rate the importance of this paper
==> 3: Important

Q2-B. Comment on importance
==> This is an important paper because it explores how feedforward processing in the visual system interacts with cue-related feedback. The authors present a modelling approach that is novel but at the same time well connected to recent work by other groups.

Q3. Rate the technical rigor of this paper. Are the results or claims of the paper supported by convincing evidence?
==> 3: Convincing

Q3-B. Comment on technical rigor
==> The empirical work involves the design and implementation of artificial neural networks. The networks have a good trade-off between complexity and feasibility,

including a combination of networks with one hidden layer

Q4. Rate the clarity of presentation. Are the ideas clearly communicated?
==> 4: Well-written

Q4-B. Comment on clarity of presentation
==> The paper is well written and understandable. As a small comment, I am not convinced by the use of the concept "representational capacity" and the manipulation of this concept by the amount of noise/challenges received during training. Intuitively it does not make much sense, I would have to re-read the paper each time to remind myself about what they mean with it.
REPLY: { We agree that we did not fully manage to communicate our intentions of using the term “representational capacity”. Consider a statement made in the introduction - “Intuitively, the interaction involving feedback would help with object recognition especially when the feature information required to recognise the object cannot be extracted by the object processing stream”. The use of the word “represented” would have been better suited than “extracted”. The idea is that due to downstream task demands or being exposed to a subset of the input space, the object processing stream might not be able to represent all task-relevant features. Representational capacity is a measure of how extensively the task-relevant input-output mapping can be realised through the object processing stream. This makes it a function of neural capacity too. But given a neural capacity, the representational capacity can be varied. We motivate the concept in the introduction as follows - “For example, if the stream is trained to represent one object, it will not perform well if two objects are presented in the same image unless feature selection is employed in the early stages of the network. We term this the representational capacity limit.” }

Q5. In your opinion, how interdisciplinary is the work in this paper? To what extent does the work integrate neuroscience measurements; sophisticated computational techniques, ideas, and models; and tackle rich cognitive theories and phenomena?
==> 4: High

Q6. Overall evaluation of paper
==> 4: Very good