We thank the reviewers for their constructive feedback and their appreciation of our experiments on this paper. The suggestions about adding more model variations are helpful; we will include more model variations (Alexnet, ResNet, and different initializations) in the revision. We have already performed comparisons of many architectures with the mouse data and achieve qualitatively similar results. We address some of the specific comments below.

Reviewer 1

novelty of findings We appreciate the reviewer's comments about the pseudo-depth metric and criteria for its robustness as novel findings. Our aim is to add another to the list: that we have used these tools to evaluate the *complexity* of visual processing in mouse cortex. Our analysis with VGG networks in the submission shows that even the primary area VISp (also called V1) computes substantially higher-order features than are commonly assumed in biological models of visual processing (such as HMAX); in our revision we will, as the reviewer rightly requests, evaluate this with shallower networks as well.

Reviewer 2

compare to mouse V1 VISp is an alternative nomenclature for V1, mentioned (we admit, briefly) in the text (L70). We will emphasize it more in the revision. This confusion coincidentally illustrates that "mouse visual cortical areas are relatively high order representations (including VISp) in a broad, more parallel organization" is, in fact, a surprising result. Our results show that even the earliest stage of the mouse visual cortex computes more higher-order features than is commonly assumed. This is consistent with a growing body of literature.

shaded error bar These are standard deviation. We will clarify this in the revision.

scientific background Functionally, very little is known about the visual areas in mice; we have nowhere near the level of detail we have about the primate cortex. There is some knowledge from anatomy (Harris, et al, 2019). There is some evidence for functional specialization in terms of spatial and temporal frequency processing (Andermann, et al, 2002, Marshel, et al 2011). In the absence of such information, the VGG16 pseudo-length gives us a window into the functional organization of the higher visual areas, one that is roughly consistent with the relatively flat hierarhcy observed in Harris, et al. We will add this discussion in the revision.

SSM vs SVCCA They are two different metrics that have different properties. SVCCA is invariant to affine transformation on the features. SSM is invariant to monotonic transformation on the similarity matrices. We would therefore advocate using either depending upon the precise question. We will add this discussions in the revision.

NeurIPS audience This work provides a robust method that reveals the functional organization of a biological visual system (mouse) whose neural coding properties are currently – and relatively recently – the subject of very intense study across computational neuroscience. This is important, because data are available for the mouse on a totally unprecendented scale – which, as we show here, enables new questions to be asked. It is also important because this system appears to show, even in its early areas, distinctly higher-order coding properties when compared with the standard view of, say, the macaque ventral stream. Thus, our work provides new inroads at the interface between engineered and biological computing networks that have long been a mainstay of NeurIPS.

Kornblith et. al. ICML 2019. Thanks for pointing out this reference. Kornblith et. al. discussed the properties of the similarity metrics on comparing artificial neural networks. Our paper focuses on robustly studying systems in which one does not have access to all units, as at present must be the case for biological systems such as mouse visual cortex. We will add a citation in the revision.

Reviewer 3

2-photon data is usually deconvolved The data is deconvolved by the algorithm in (Jewell et. al. 2018). Full information about how the data was processed is given in the Allen Institute paper de Vries, et al 2018. We will provide more information about the data in the revision.

deconvolution is unable to identify scaling SVCCA is invariant to affine transformation on the features, thus is also not affected by the scaling in the data.

trial-to-trial variability and non-stationarity This is a very important question. We will add results of bootstrapping trials to quantify the effects of trial-to-trial variability and non-stationarity among trials in the revision.

why not use exactly those stimuli The primary reason we use images other than those used for the Brain Observatory are so that we can study the sampling trends beyond 118 images (the number shown in the experiment). For the comparisons that don't require this, we do use the exact stimuli (such as when comparing to the mouse data), including for the neural subsampling curves.