We thank the reviewers for the thoughtful comments. We will address ALL comments in our final version.

Reviewer 1: We do not claim that we improve a standard model for image classification. We claim that in the context of active learning, the proposed method achieves enhanced speedup rates (budget/accuracy trade-off) compared to working with a fixed architecture along the active learning process. We also state that selecting the baseline architecture requires hindsight knowledge over the learning problem at hand, that might not be available when actively learning on new problems and domains. Thus, instead of selecting a fixed architecture (based on an educated guess) we recommend to build an architecture search space (similar to the one proposed for resnet), and run active iNAS. This will allow to actively learn faster with less labels and reach better results compared to the (guessed) fixed architecture. Of course, for any particular dataset it's liklely that the SOTA passive architecture achieves better final result than our final architecture. However, our sample complexity speedup performance is most likely to be superior. And once you have a full labeled dataset you can take the SOTA architecture anyway. For better understanding of this point please see comparison to other active learning papers described in the supplementary material.

Reviewer 2: AL vs NAS - The paper discusses an enhanced active learning method that incorporate a NAS algorithm that is designed to support the active learning process. We will move some of the literature survey of NAS to the main text as you proposed. The active i-NAS (compared to standard NAS methods) leverages the properties of active learning process, where the training set starts small and grows along the process. Our algorithm deliberately designed this way, and in particular we avoid application of a full NAS on each active learning round which is both statistically prone to overfitting and computationally infeasible.

Weights sharing across active rounds - We train the network from scratch on each active learning round. A future research direction would be to integrate weight sharing ideas such as those propose in ENAS [1].

how reliable is the architecture selection in the early stage of AL - The variance of early stage active learning is known to be high due to variability in seed initialization. In our case we can see that the performance is relatively stable among repetitions (very small shadowed standard error in the curves). In most cases, Active iNAS reached the same architectures (both on the early and later stage of the active learning). The latter result is not presented in the paper and will be added.

It might not always be the case that more data induce deeper (more complicated) model ... - We think that this situation is statistically rare in i.i.d selection of the initial seed, and does not happen in the later stages of active learning where points to be labeled are selected to increase diversity and improve model accuracy on uncertain cases.

Reviewer 3: Large budget compared to classic AL - typical deep neural models have huge hypothesis class capacity, which make them hungry for labeled training sets. In this setting it does not make sense to use the budgets that have been used in classic AL (5-300). Please see the references to deep active learning papers as it is a common numbers/sizes for deep learning AL.

Insights about why your approach remains better compared to the fixed networks - the improvement comes from both being more adapted to the training set size, and due to enhanced querying. The improvement in querying performance, is demonstrated in Section 4.4. In the experimental results, we can see that the initial architecture selected by iNAS outperforms the (large) fixed architecture in the early active learning stage in most cases. This supports the claim that the architectures that have been selected are more suitable to the training set size along the active learning process. We will emphasize this point in the text.

Lack of comparisons with other state-of-the-art approaches/architectures for AL - We compared all methods for active learning (Coreset, softmax and MC-dropout), on our experimental setting and architectures. Please see a direct comparison to the state-of-the-art papers for deep active learning in the supplementary material. In addition, please see the response for reviewer 1 about state of the art architecture selection in hindsight.

Random querying baseline – we agree that random querying (i.e. passive learning) is a standard baseline that any good AL algorithm should surpass. Our algorithm easily beats passive learning in all cases. We will add the learning curves. You can also see some evidence that we are better than passive learning by considering the results of our other baselines (in their original papers). For example, Sener & Savarese and Geifman & El-Yaniv.

Comparison to cross-validation over several fixed architectures - The active iNAS allows us to actively (wisely) search in a large architectural search space (60 architectures in our case). Searching exhaustively in this search space is computational in-feasible, and prone to over-fitting (due to the small validation set in early stages). Thus, we proposed the iNAS algorithm that leverages the incremental expansion of the training set size in active learning and expand the selected architecture accordingly in much efficient way compared to exhaustive cross-validation.

[1] Pham, Hieu, et al. "Efficient neural architecture search via parameter sharing." arXiv preprint arXiv:1802.03268 (2018).