This paper is well written, the organization, style and structure is good. The presentation is easy to follow. The major contribution of the paper is to present a general purpose approach for improving sample efficiency for transfer learning using two novel approaches that leverage direction and magnitude of gradient vectors generated by a weakly supervised learning loss to rank images required for human annotations. The efficacy of the approach is illustrated by experiments on two datasets that require multiclass semantic segmentation on RGB images.

Major Comments :-

1) The proposed approach relies highly on how good are the gradients generated weakly supervised segmentation loss, yet there is no baseline on what happens if the authors train with a large number of images using this loss alone and then finetune with a small set of randomly selected human annotated images.

2) The authors in background work section describe a number of already available approaches to solve the same problem, yet there is neither any experimental comparison nor an intuitive argument as to why the proposed approach is better than these methods.

3) Incase of Gradient Projection approach, the norm would be high is case of both really dissimilar or similar gradient directions, whereas the authors would intuitively only want the former case. It would be good to see some clarifications regarding this.

4) There also seems like there a clear correlation between Gradient Projection approach and cosine similarity of gradients from auxiliary tasks approach proposed by Du et. al. [1]. It would be good to see some discussion on this.

5) In the results section authors write :-
“The plots illustrate that both approaches achieve diversity despite the fact that the strongest gradients cluster together, near the bottom left”.

   This is a broad conclusion which is not obvious to me looking at the Figure 6 alone. Please plot the data points selected by “Uncertainty” and “random” method too, in order to make a stronger case for this argument.

Minor Comments :-

6) Please write a more descriptive caption for figure 3. Is Fig 3(c) segmentation mask generated from the proposed approach or weakly supervised approach ?

7) There is no Figure that shows raw image, human annotation, weakly supervised segmentation mask and segmentation mask from the proposed.

8) In both figures 4 and 5 we don’t see a baseline datapoint of when there is 0 samples used for fine tuning.

This paper presents what authors' want to deliver. Both the results and the technical procedures are well delimited and explained. Overall this is a nice paper and should be considered as IROS paper after necessary changes suggested by all the reviewers.
Regarding the references used, I find them appropriate and current, which shows that research is a topic of current interest.

References :-

[1] Du, Yunshu, et al. "Adapting auxiliary losses using gradient similarity." arXiv preprint arXiv:1812.02224 (2018).