

# View Reviews

## Paper ID

2

## Paper Title

ViS-HuD: Using Visual Saliency to Improve Human Detection with Convolutional Neural Networks

## REVIEWER #3

---

### REVIEW QUESTIONS

---

#### 1. Relevance to MBCC

Relevant

#### 2. References to prior work

Adequate review of relevant literature

#### 3. Technical correctness

Appears to be correct

#### 4. Experimental validation

Limited experiments and/or weak results (please explain in comments)

#### 5. Presentation quality

Unclear

#### 6. Overall rating (5 = excellent, 1 = poor)

2

#### 7. Detailed comments to authors

This work presents a method to improve people detection by CNNs using saliency maps. Overall I would recommend a weak accept -- the approach is interesting and seems to be an improvement over some existing methods, but the experiments seem limited and some sections are unclear.

Section 2.3 seems unnecessary -- several deep CNN methods were discussed in the previous two sections, so including a separate explanation of these methods seems out of place. It would be better to incorporate these references into the earlier sections, or use this section to discuss methods more similar to the current approach (e.g. methods which use layer activations from a pretrained GoogLeNet / AlexNet / VGG-16 / etc. as "features" for pedestrian detection or similar tasks).

The fact that Cornia outperforms other models should be stated earlier to make the motivation for using this model clearer. (Currently this is stated in 3.2, but it could be explained when the model is introduced in 2.1.)

The coverage map is not thoroughly explained -- how it is computed and its purpose in the loss function should be more clearly detailed.

It seems like the model was trained on both datasets, but a stronger test of the model's utility would be to train on one database and test on the other, or to show additional tests on a third, unrelated image database.

It is not clear if any other deep CNN methods were included in the comparisons (Table 2 does not include CNN-based methods, and I'm not sure if any are included in Figure 8).

The experiments should include a comparison of the results with and without the saliency model, to show the impact of

including saliency. It would also be interesting to compare results using a “baseline” saliency model consisting of a central Gaussian, to show that the saliency model is not simply biasing the detector towards the center of the image.

There are some grammatical errors in the paper which make some parts unclear.

---