

## Rebuttal for ‘Guided Proofreading of Automatic Segmentations for Connectomics’

Thank you for your time and feedback.

**R2: Venue** R2 writes “While solid work, the paper is not a perfect fit for CVPR”. Biological and Cell Microscopy Image Analysis is included in the call for papers of CVPR 2018, and the field of connectomics has previously given rise to interesting papers at CVPR (Kumar *et al.* 2010, Kaynig *et al.* 2010, Jain *et al.* 2010, Funke *et al.* 2012, Pape *et al.* 2017).

**R3: Generalization to other segmentation problems** R3 misses a discussion regarding applicability of GP to general semantic segmentation problems. We believe our method will also be interesting to researchers working on tasks beyond connectomics, as segmentation proofreading for labeled dataset collection and correction is widely applicable in computer vision. We state mandatory re-training of GP for other datasets as a limitation in the supplemental material (lines 134-138) but will further elaborate on general segmentation problems in this section.

### R2: Algorithmic Contribution/Innovation

### R2: Superhuman Accuracy on the SNEMI3D Challenge [1]

### R2: Variation of Information Metric [2]

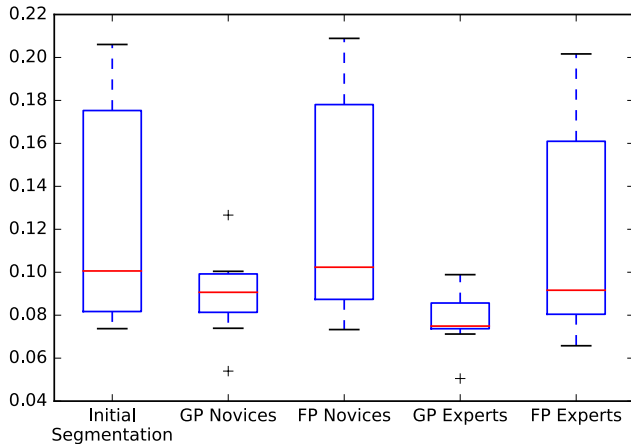


Figure 1: Adapted Rand Error distributions of novices and experts using guided proofreading (GP) and focused proofreading (FP) as part of the forced choice user experiment across slices of the AC4 subvolume.

**R2: Generalization of the AC4 Subvolume** R2 questions the generalization ability of the AC4 subvolume due to small dimensions. We agree that this dataset is small. However, it was introduced by Haehn *et al.* 2014 for feasible interactive proofreading studies and is representative for the full AC4 dataset with respect to the distribution of object sizes.

**R3: U-Net training data** R3 requests further information regarding the training data of the U-Net membrane probability classifier. Section 2 of the supplemental material includes this information (supplemental material lines 140-161, Table 3). We will modify the manuscript (lines 492-493) to include a direct reference.

**R3+R4: Network input channels** R3 is interested in seeing the performance contribution of the different input channels. R4 suggests that the membrane probabilities and the dilated border mask provide more discriminative information than the other two channels. We observe that all four input channels are important to reduce VI (Table 1). Similar as identified by Bogovich *et al.*, image data adds intracellular structures (e.g. vesicles) to the decision process and membrane probabilities include global knowledge of the staining protocol to highlight cell membranes. We then notice, that the label channel provides knowledge about neuron shapes while the border mask covers the gap of extracellular space.

Table 1: We evaluate automatic selection on the AC4 subvolume ( $p_t = 0.95$ ) using the GP classifier with different input channels and report median VI reduction. The combination of all four channels performs best.

Input channels	VI reduction
<i>Image + Prob.</i>	-0.094
<i>Image + Prob. + Border</i>	-0.045
<i>Image + Prob. + Label</i>	0.038
<i>Image + Prob. + Label + Border</i>	0.065

**R4: Dilated mask of the border between two segments** R4 raises the question of whether the dilated mask of the border is a crucial input to our classifier.

The dilated border improves classifier performance measured as VI reduction from 0.038 to 0.065 (Table 1). We qualitatively describe the intention behind the border mask

108 in lines 315-323 but will add the experiment above to the  
109 supplemental material and to the open source repository.  
110

111 **R4: 2D Slices only**  
112

113 **R4: Merge error detection performance** R4 expresses  
114 concerns regarding the performance contribution of merge  
115 error correction.  
116

117 **References**  
118

119 [1] K. Lee, J. Zung, P. Li, V. Jain, and H. S. Seung. Superhuman  
120 accuracy on the SNEMI3D connectomics challenge. *CoRR*,  
121 abs/1706.00120, 2017. 1  
122 [2] J. Nunez-Iglesias, R. Kennedy, T. Parag, J. Shi, and D. B.  
123 Chklovskii. Machine learning of hierarchical clustering to  
124 segment 2D and 3D images. *PLoS ONE*, 8(8):e71715+, 2013.  
125 1  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161

162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215