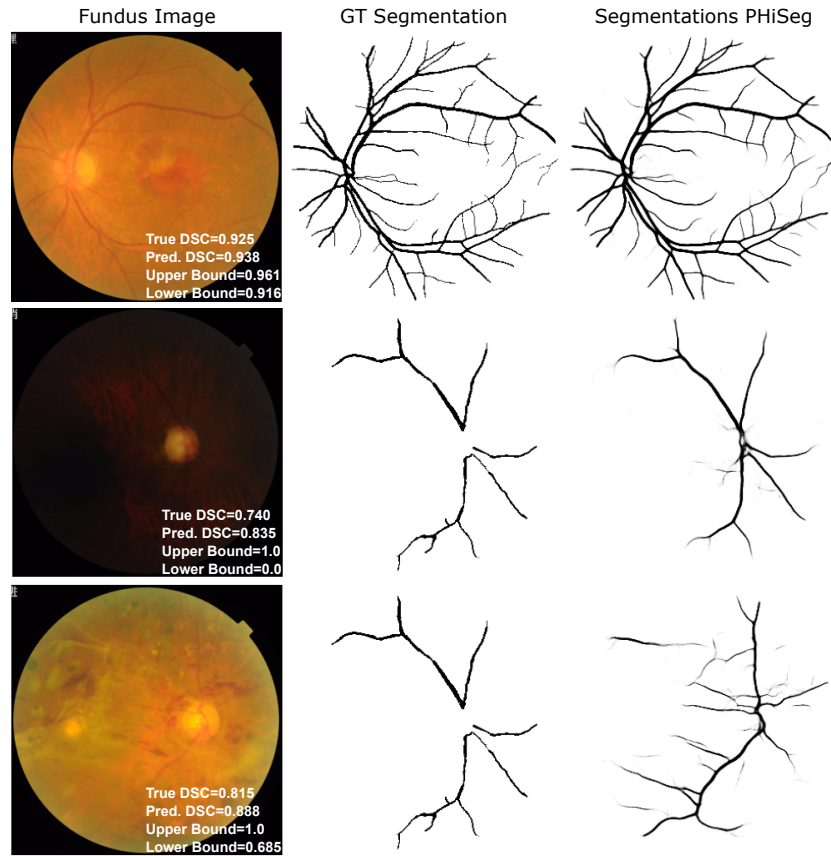# 1 Qualitative examples

Fig. 1. PHiSeg segmentations and performance predictions for three examples. Top: An example with good segmentation performance. Middle: An example with poor image quality and poor segmentation performance. Bottom: An example with good image quality but poor segmentation performance due to visible pathologies stemming from advanced diabetic retinopathy.

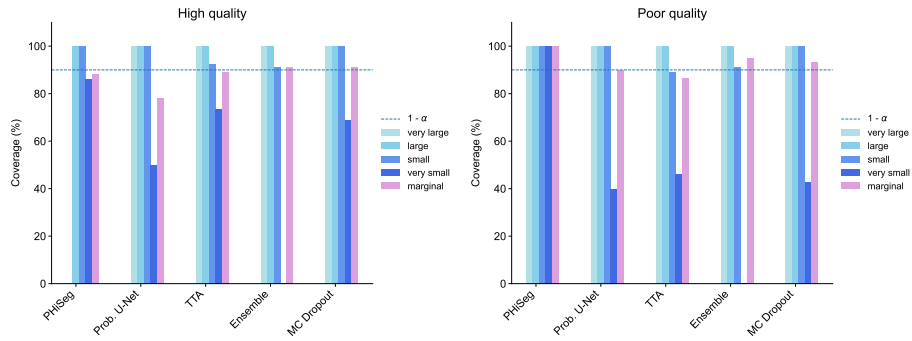## 2 Conditional coverage grouped by quality

Fig. 2. Marginal and conditional coverage for multiple interval sizes: very small (0, 0.1], small (0.1, 2], large (0.2, 5]), very large (0.5, 1]) for high quality images (left) and poor quality images (right).