

Detection



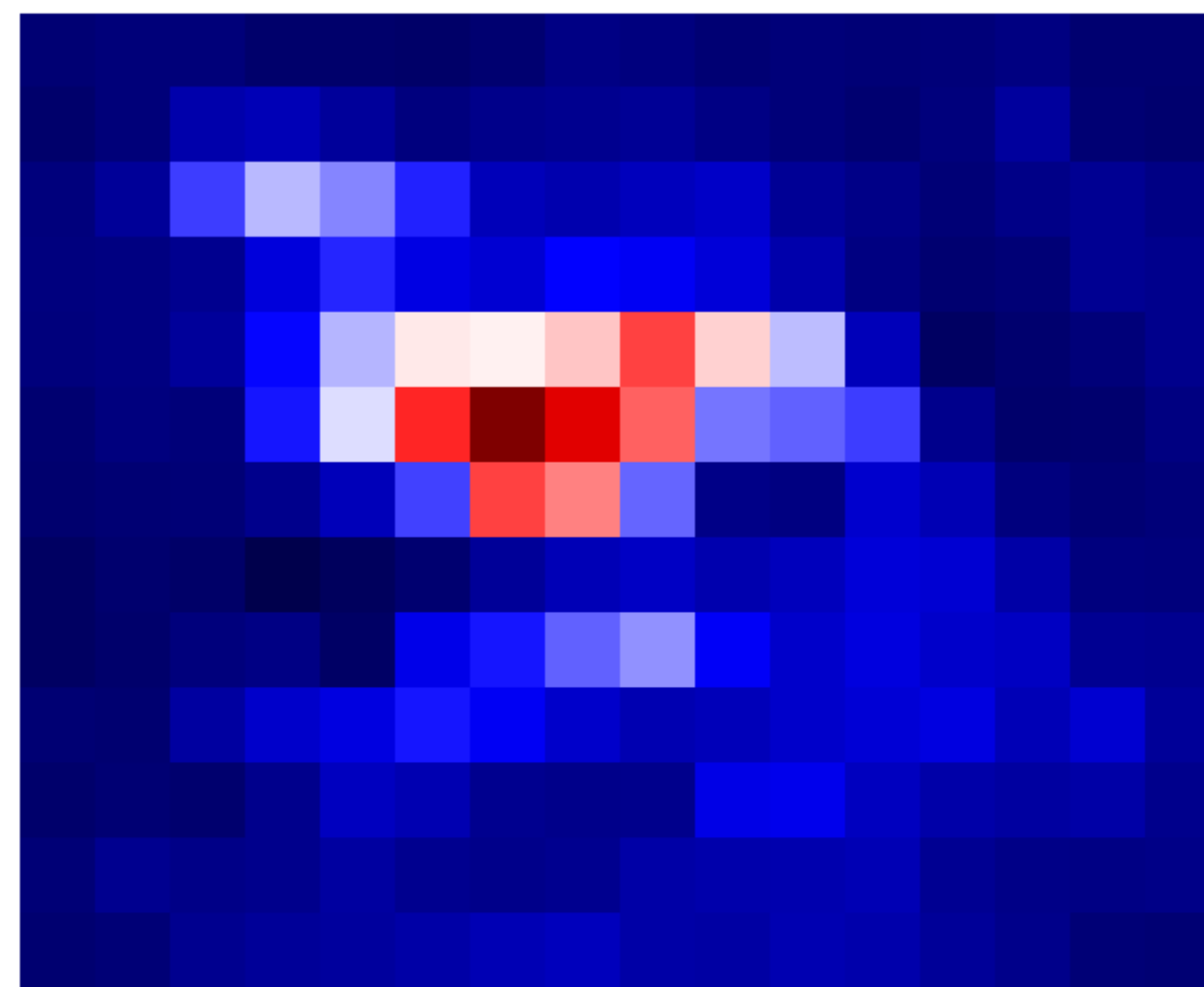
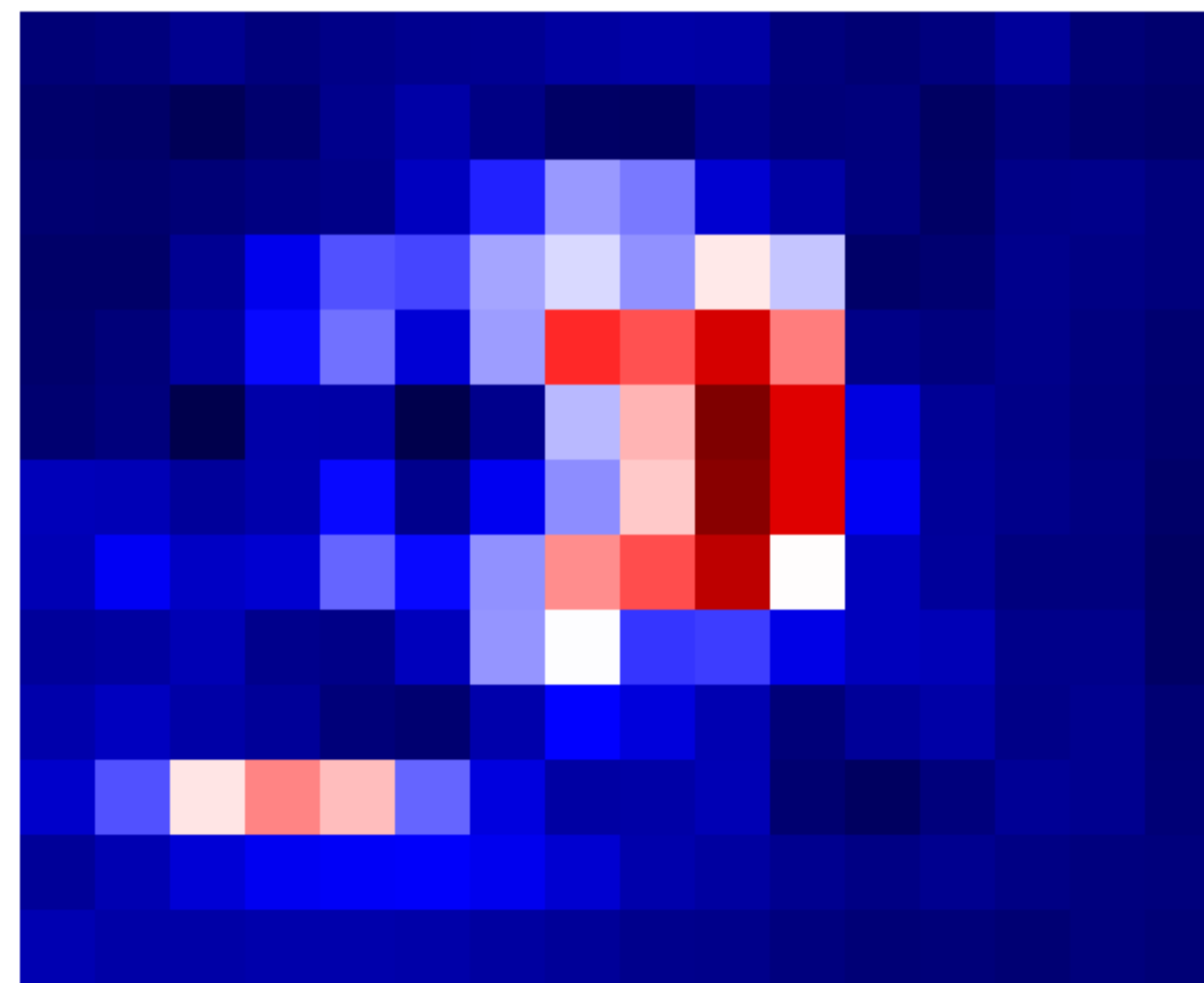
Abdominal View
(98%)

Lips View
(96%)

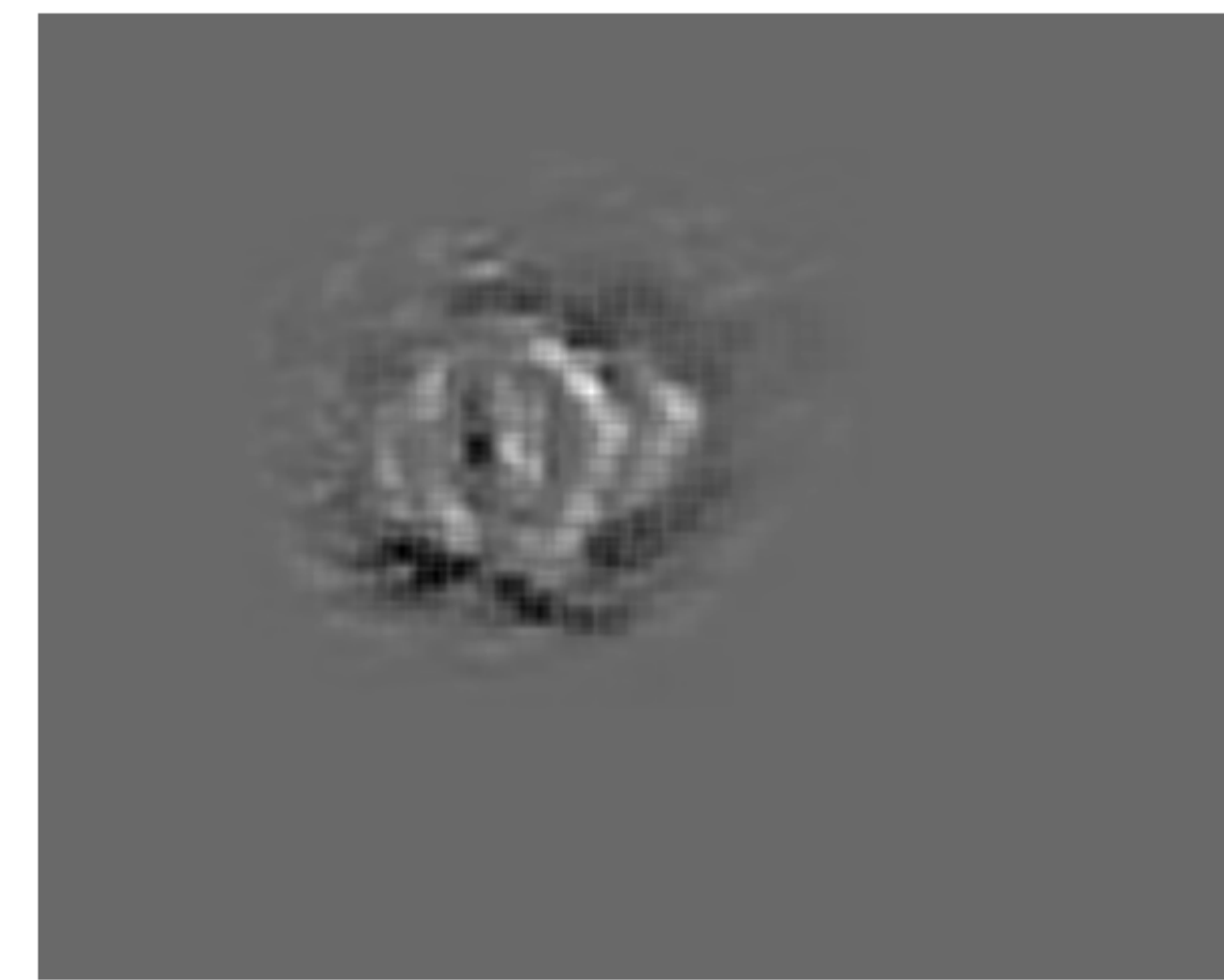
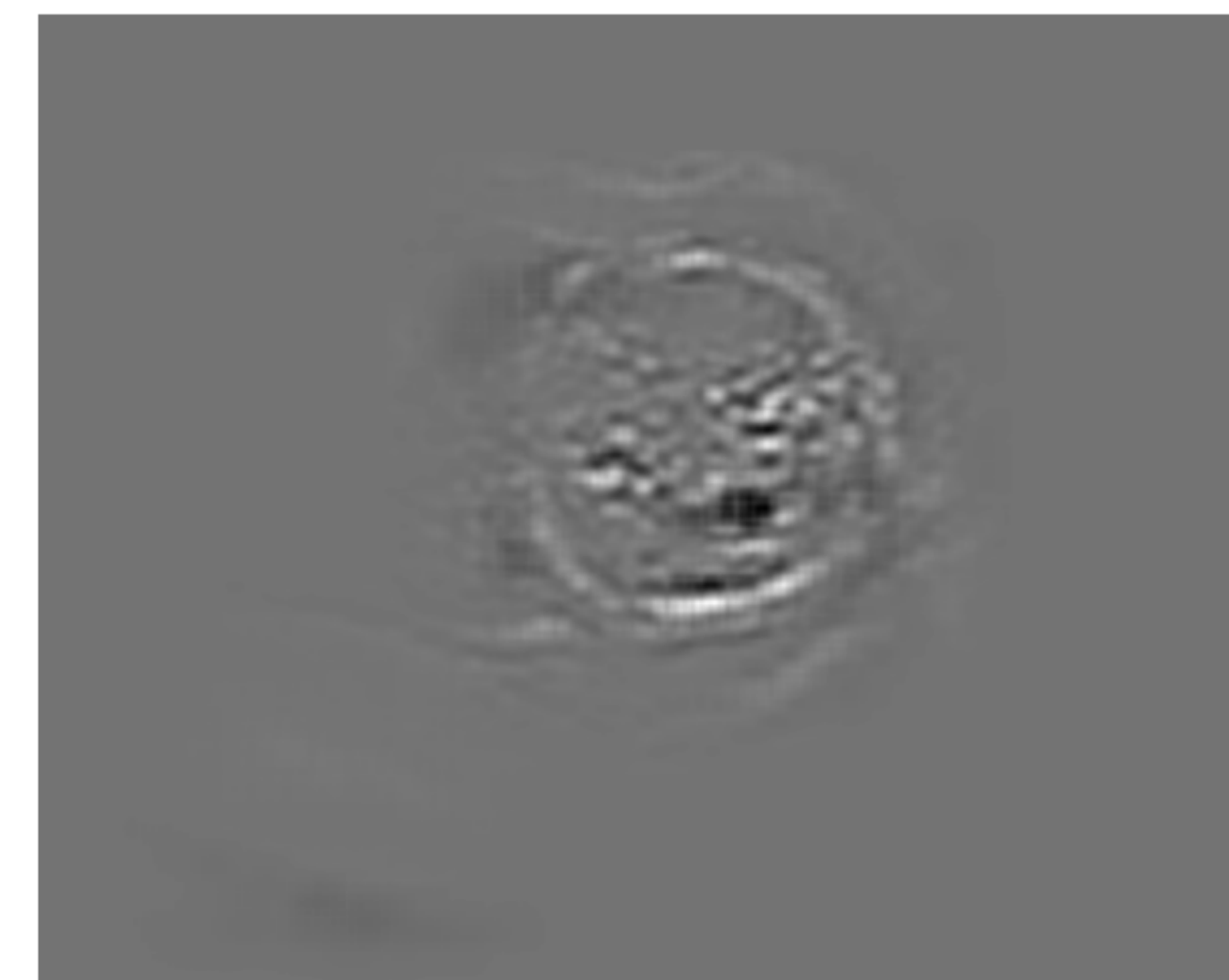
(a) Input frame

(b) Prediction

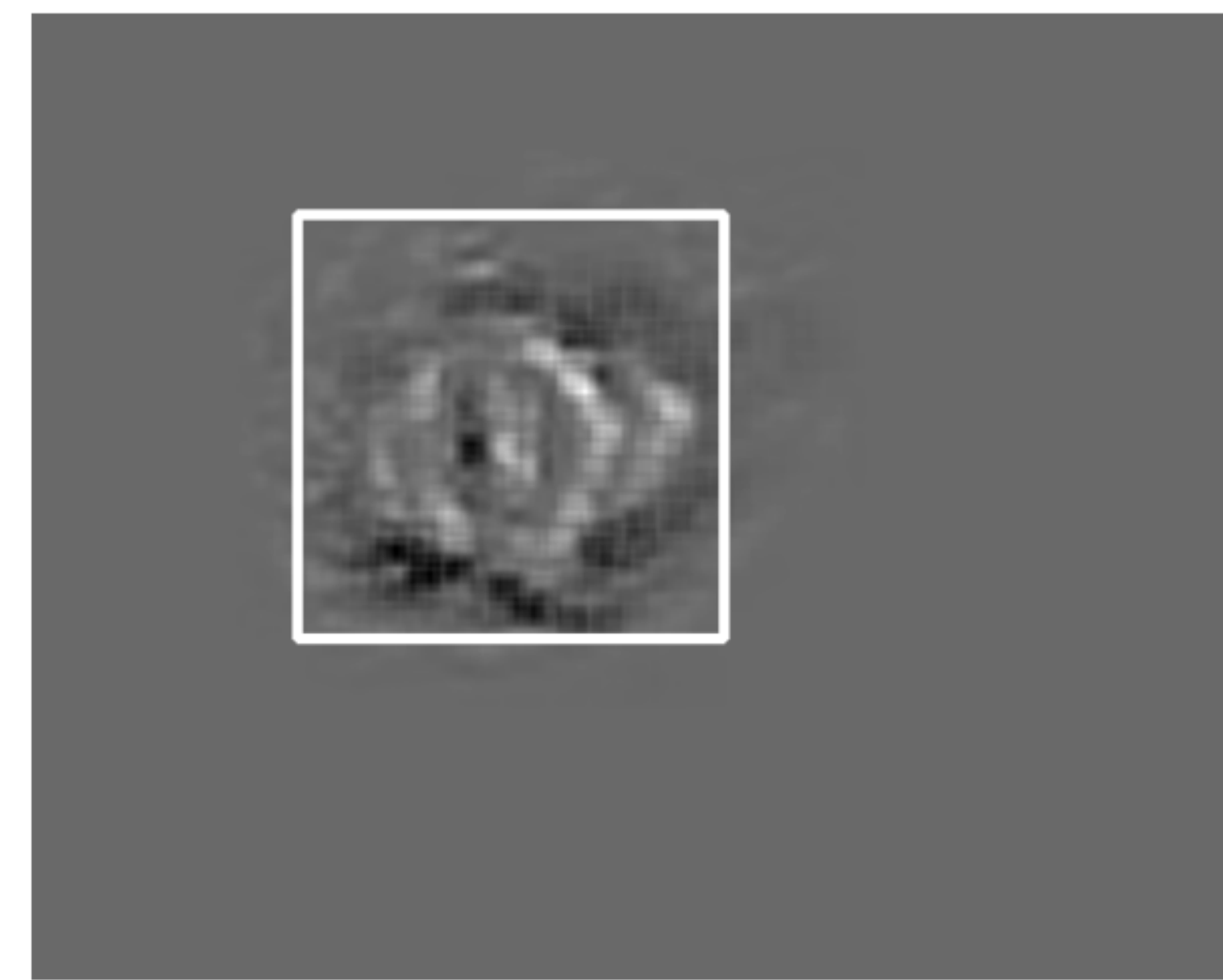
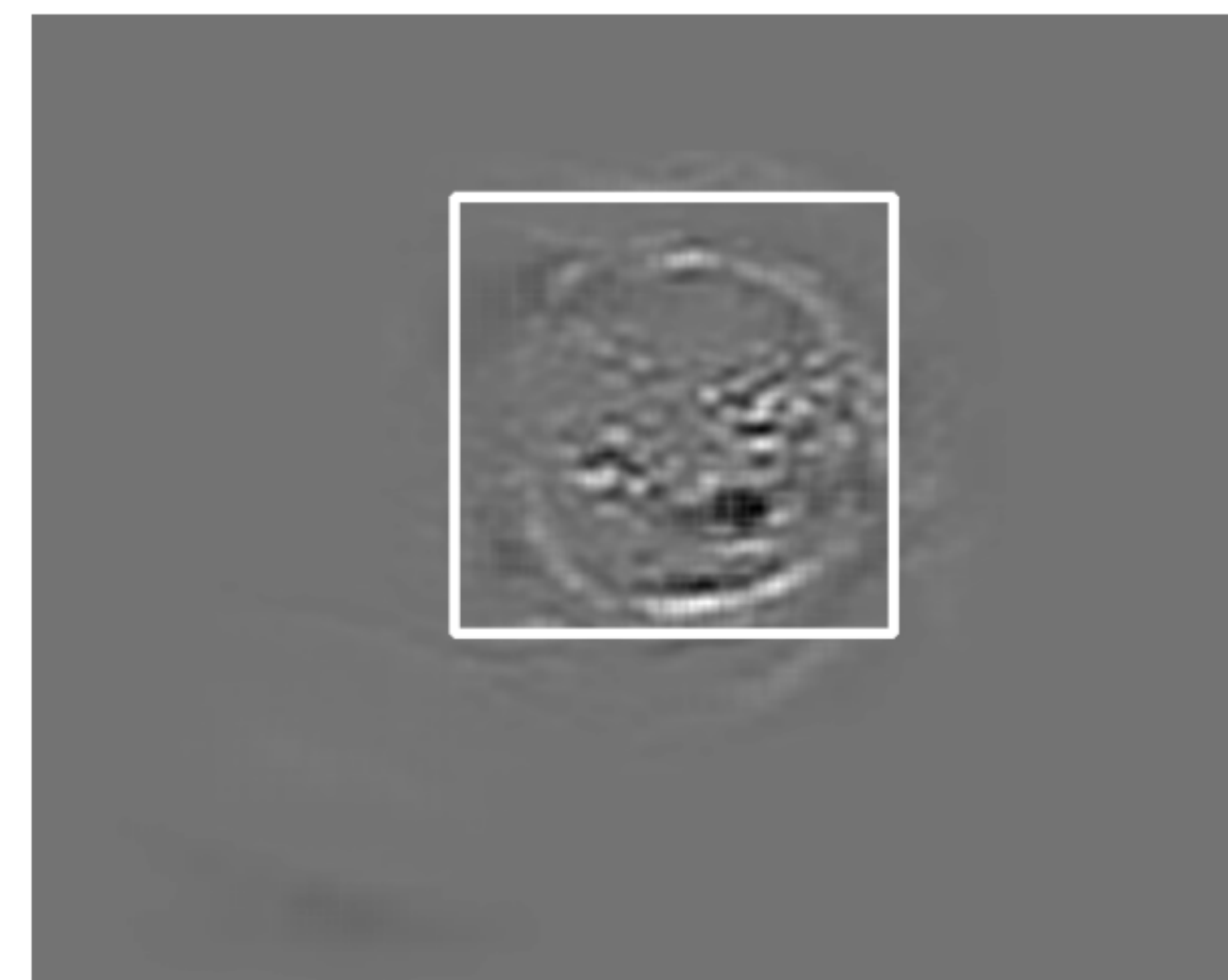
Unsupervised Localisation



(c) Category-specific
feature map



(d) Localised
saliency map



(e) Approximate
localisation