Supplementary material for "RISE: Randomized Input Sampling for Explanation of Black-box Models"

Vitali Petsiuk
vpetsiuk@bu.edu
Abir Das
dasabir@bu.edu
Kate Saenko
saenko@bu.edu

Boston University
Boston, USA

Algorithm to compute deletion score.

Algorithm 1

procedure DELETION
  Input: black box f, image I, importance map S, number of pixels N removed per step
  Output: deletion score d
  $n \leftarrow 0$
  $h_n \leftarrow f(I)$
  while I has non-zero pixels **do**
    According to S, set next N pixels in I to 0
    $n \leftarrow n+1$
    $h_n \leftarrow f(I)$
  $d \leftarrow \texttt{AreaUnderCurve}(h_i \textit{ vs. } i/n, \quad \forall i = 0, \dots n)$
  return d

Algorithm to compute insertion score.

Algorithm 2

procedure INSERTION
  Input: black box f, image I, importance map S, number of pixels N removed per step
  Output: insertion score d
  $n \leftarrow 0$
  $I' \leftarrow \texttt{Blur}(I)$
  $h_n \leftarrow f(I)$
  while $I \neq I'$ **do**
    According to S, set next N pixels in I' to corresponding pixels in I
    $n \leftarrow n+1$
    $h_n \leftarrow f(I')$
  $d \leftarrow \texttt{AreaUnderCurve}(h_i \textit{ vs. } i/n, \quad \forall i = 0, \dots n)$
  return d

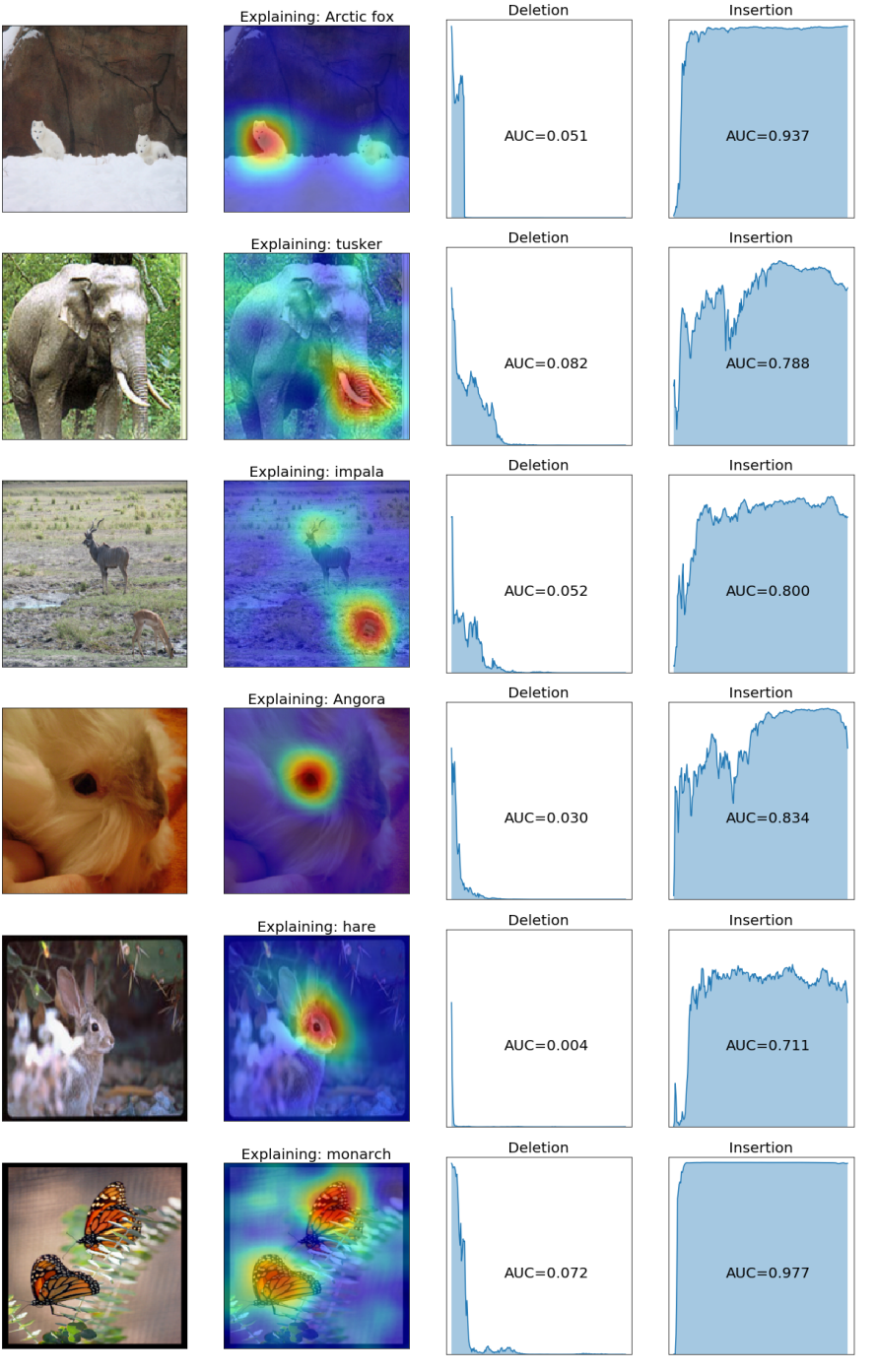

Figure 1: RISE generated importance maps (second column) for representative images (first column) with deletion (third column) and insertion curves (fourth column).

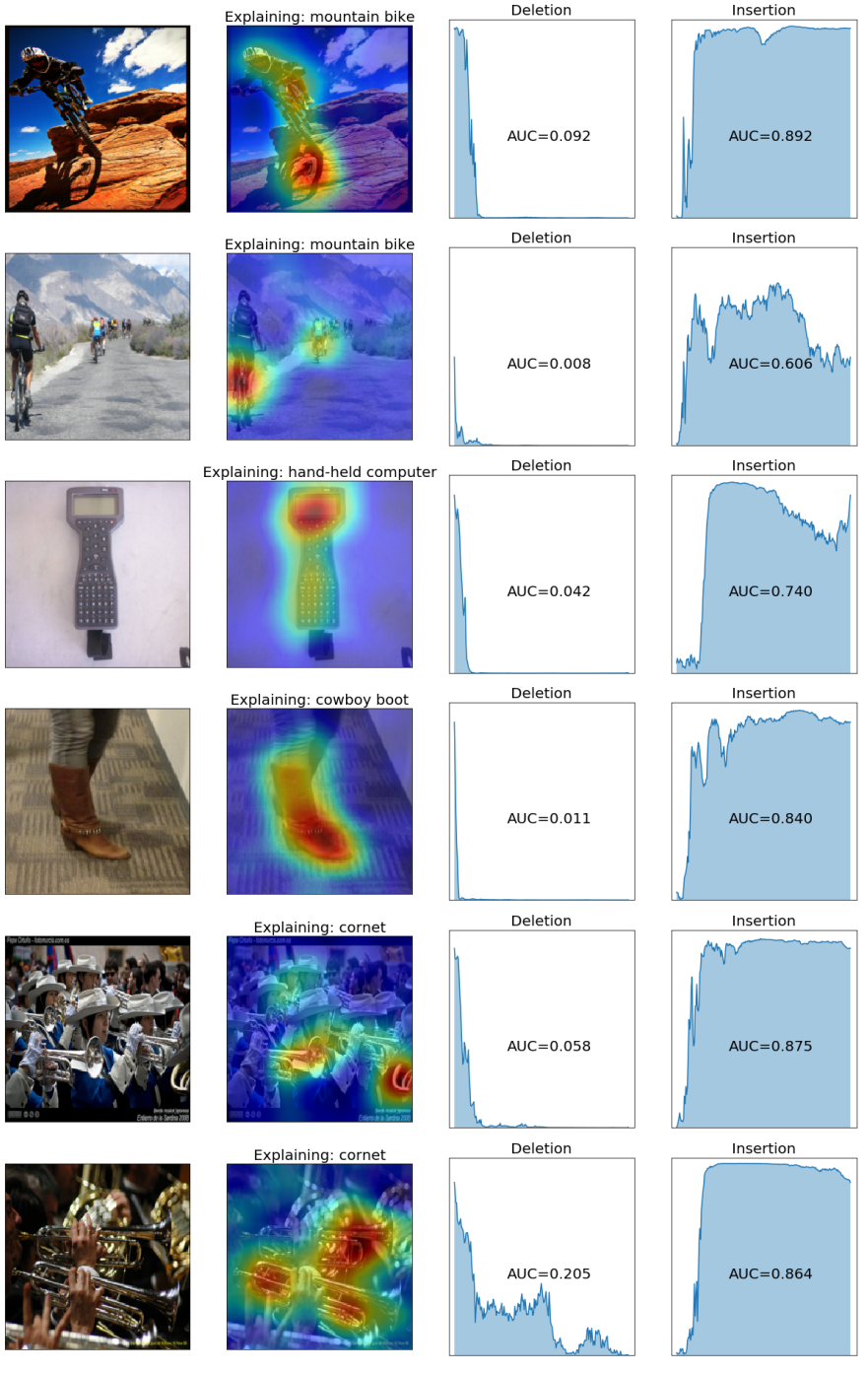

Figure 2: RISE generated importance maps (second column) for representative images (first column) with deletion (third column) and insertion curves (fourth column).

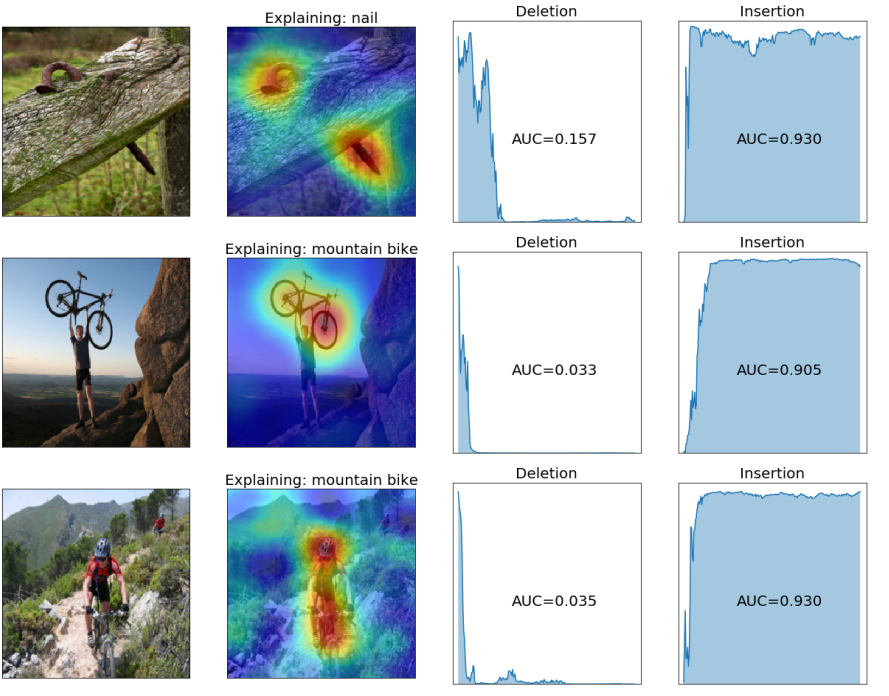

Figure 3: RISE generated importance maps (second column) for representative images (first column) with deletion (third column) and insertion curves (fourth column).

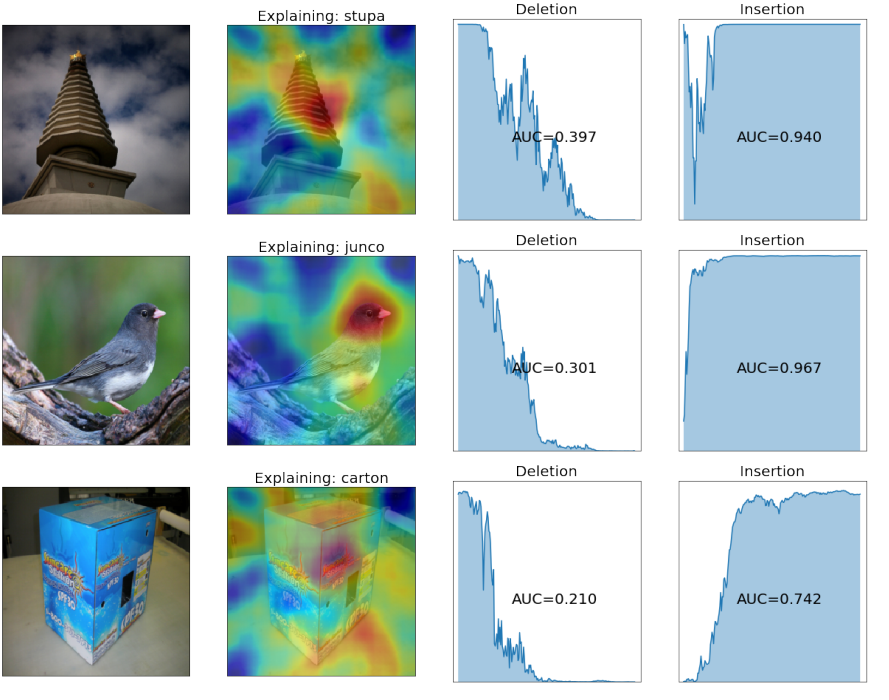

Figure 4: Failure cases. In some cases RISE does pick up more important features, but cannot get rid of the background noise (in part due to MC approximation with only a subset) like in rows 1 and 2.