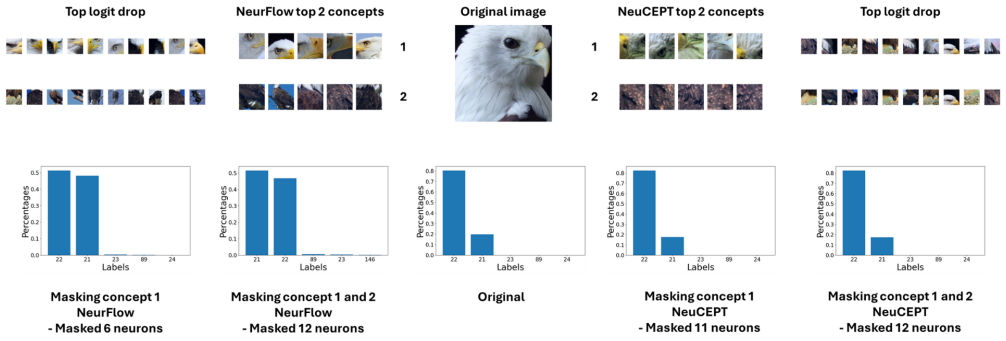

Figure 1: The comparison of the top-2 groups of neurons with the highest metric score of our method and [1] on class Bald eagle. The top logit drop images of NeurFlow are more resemble the original concept (i.e. NeurFlow concept 1 vs NeuCEPT concept 1). And the prediction probability changes when masking our critical neurons are more significant while masking fewer neurons.

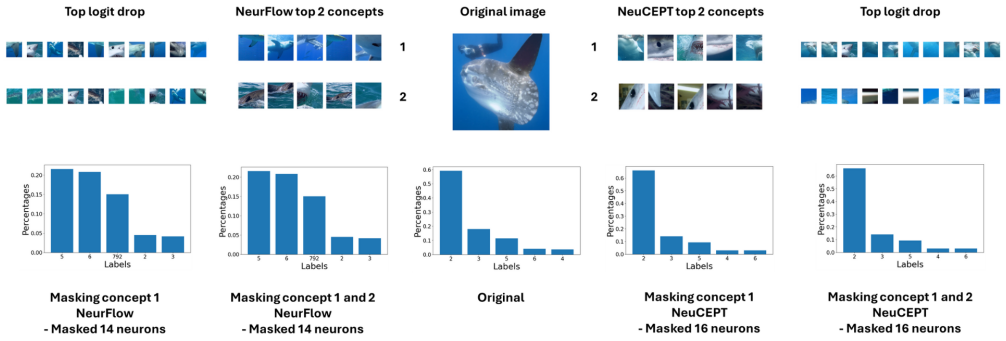

Figure 2: The comparison of the top-2 groups of neurons with the highest metric score of our method and [1] on class Great white shark. The top logit drop images of NeurFlow are more resemble the original concept (i.e. NeurFlow concept 2 vs NeuCEPT concept 2). And the prediction probability changes when masking our critical neurons are more significant while masking fewer neurons.

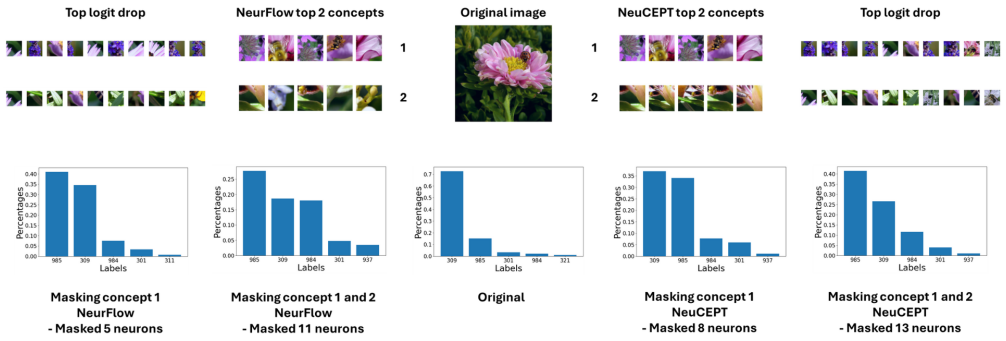

Figure 3: The comparison of the top-2 groups of neurons with the highest metric score of our method and [1] on class Bee. The top logit drop images of both methods are similar to the exemplary image of the concept. And, both methods are able to alter the prediction of the model.