

SUPPLEMENTARY MATERIAL FOR
Exemplary Natural Images Explain CNN Activations Better than State-of-the-Art Feature Visualization

1 EXTENDED: BY-FEATURE-MAP ANALYSIS

On the following pages, we provide more images of trials that participant two and three saw during Experiment I.

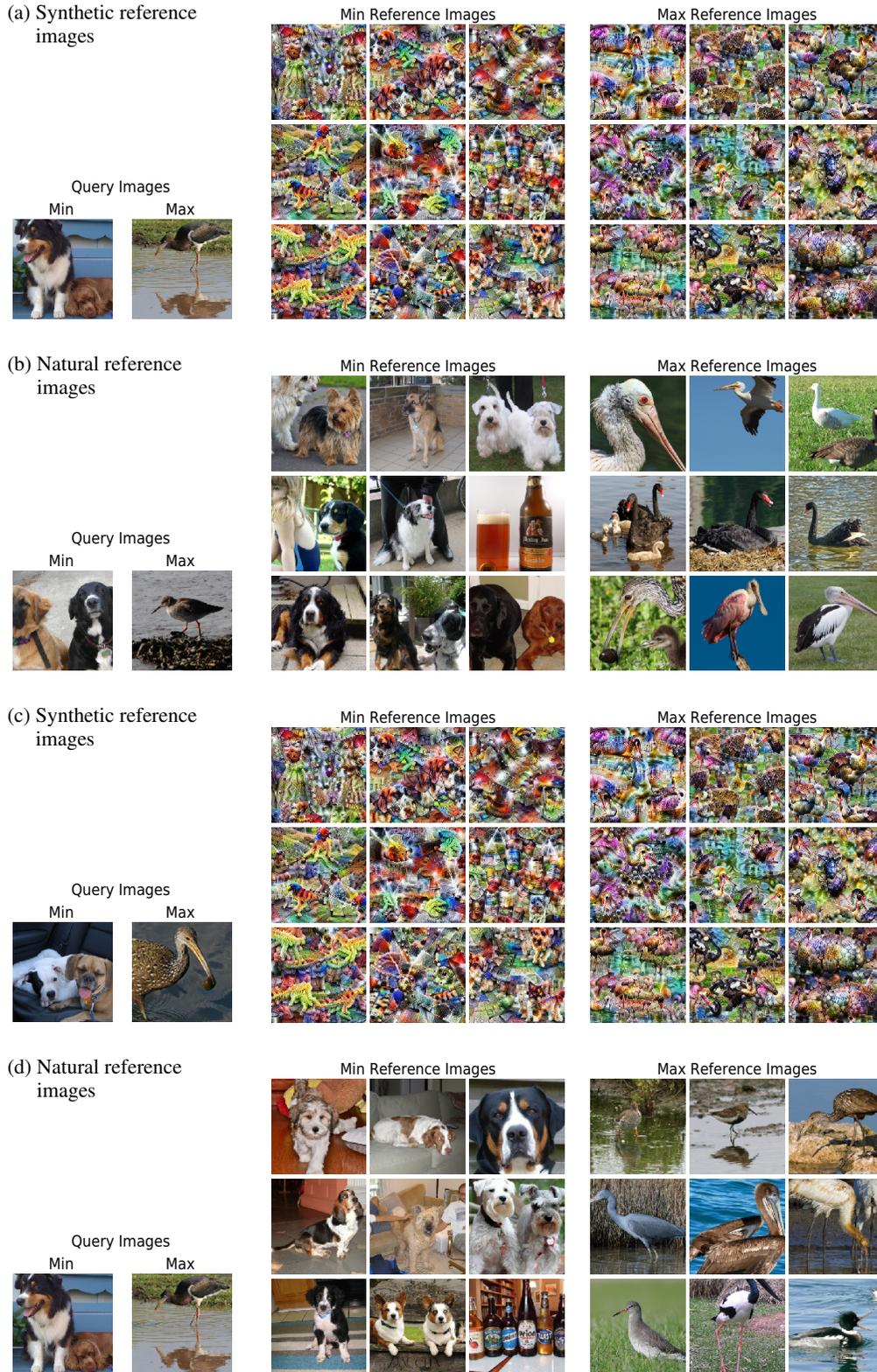


Figure 1: An easy feature map (here: 5a, pool*) from Experiment I where all subjects answered correctly for both synthetic and natural reference images. Our impression is that the decisive feature (dog vs. bird) is well understandable. The shown stimuli were shown to participants two (a and b) and three (c and d); for stimuli shown to participant one, see Appendix Fig. 23.

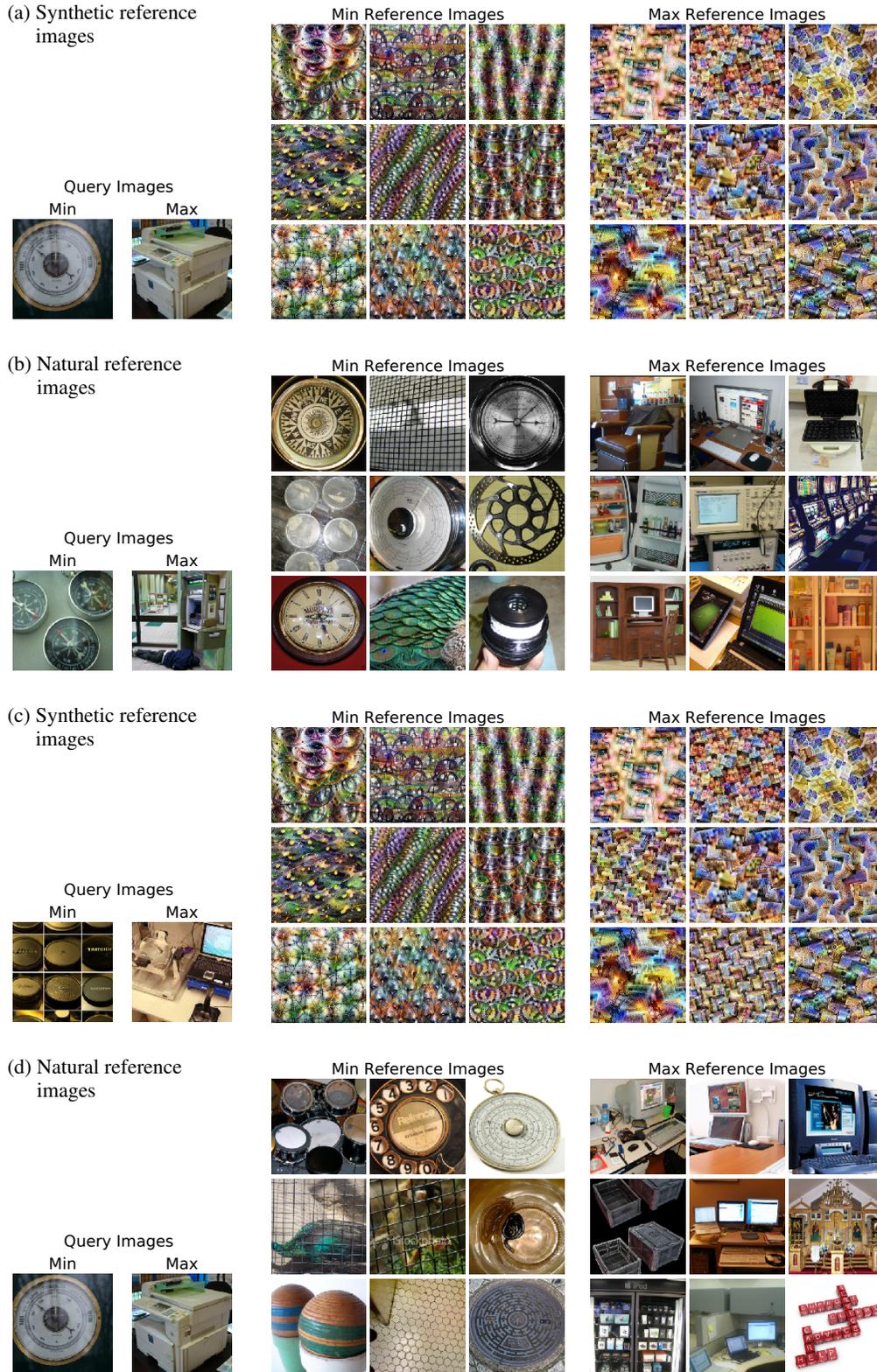


Figure 2: Another easy feature map (here: 4c, 1x1) from Experiment I where all subjects answered correctly for both synthetic and natural reference images. The decisive feature seems to be round (minimal) vs. edgy (max). The shown stimuli were shown to participants one (a and b) and two (c and d).



Figure 3: Another easy feature map (here: $4c, 1 \times 1$) from Experiment I where all subjects answered correctly for both synthetic and natural reference images. The decisive feature seems to be round (minimal) vs. edgy (max). The shown stimuli were shown to participant three.

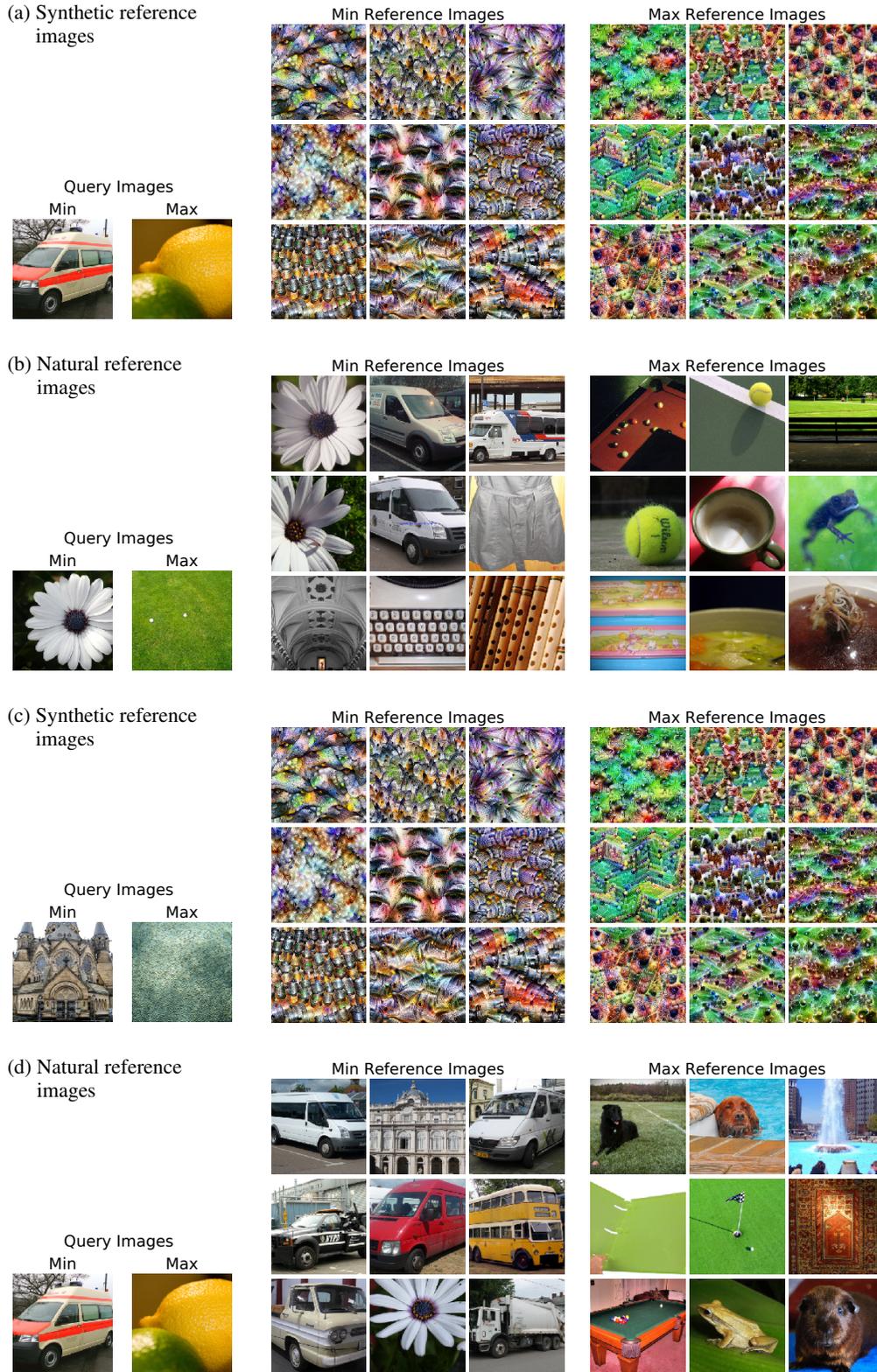


Figure 4: Another easy feature map (here: 4d, 3x3) from Experiment I where all subjects answered correctly for both synthetic and natural reference images. The decisive maximal feature seems to have to do with green color and round shapes. The shown stimuli were shown to participants one (a and b) and two (c and d).

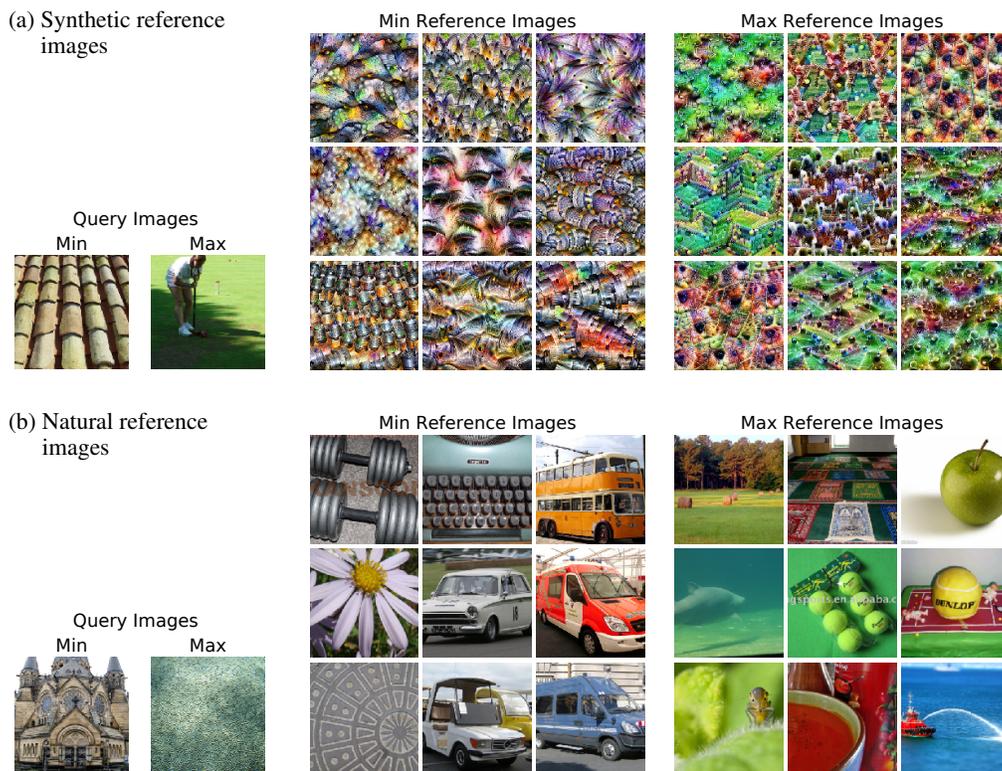


Figure 5: Another easy feature map (here: 4d, 3x3) from Experiment I where all subjects answered correctly for both synthetic and natural reference images. The decisive maximal feature seems to have to do with green color and round shapes. The shown stimuli were shown to participant three.

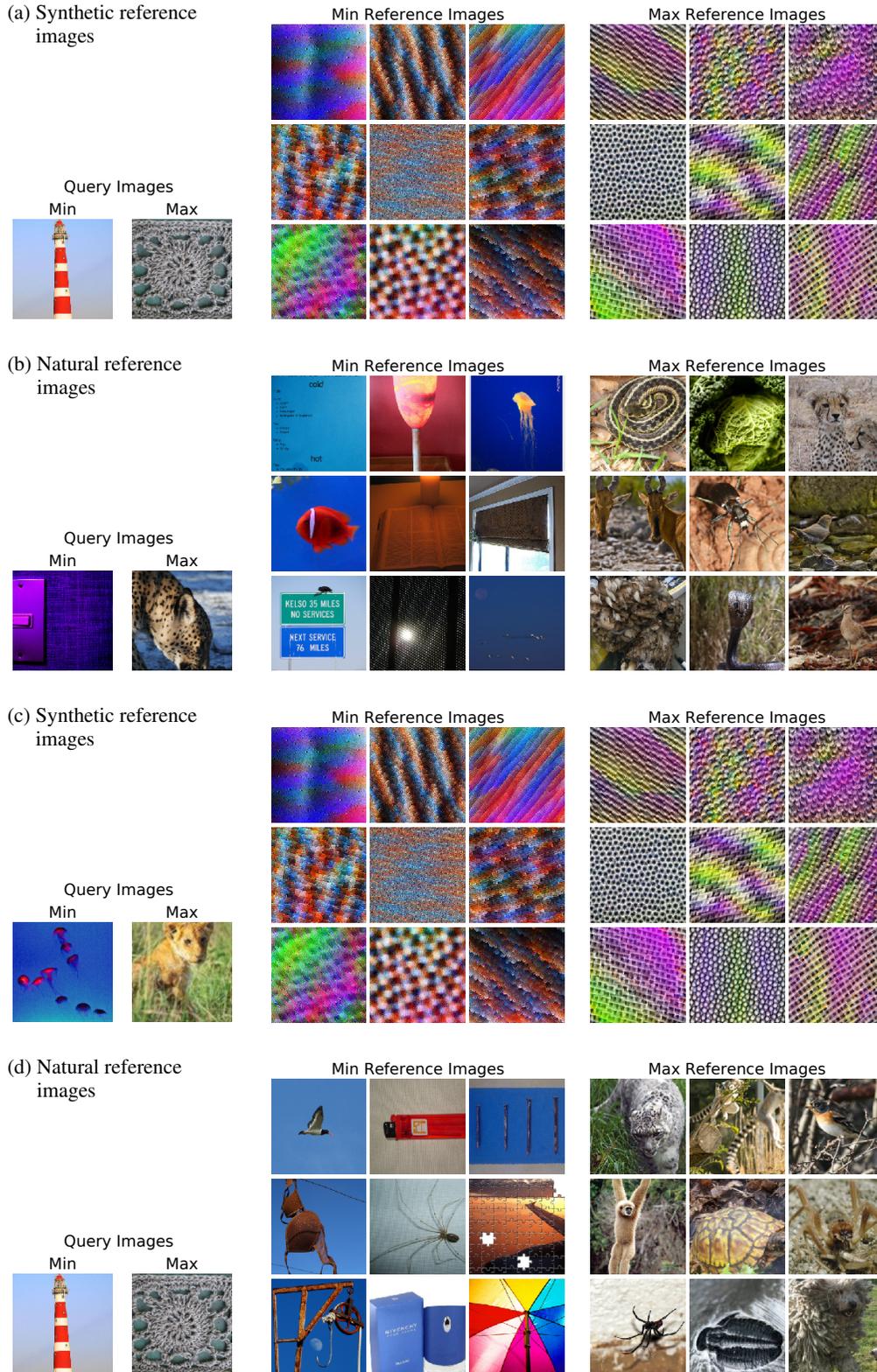


Figure 6: Another easy feature map (here: 3b, 1x1) from Experiment I where all subjects answered correctly for both synthetic and natural reference images. We are surprised that all the trials for synthetic reference images were answered correctly; to us, the decisive feature is not easily identifiable. The shown stimuli were shown to participants one (a and b) and two (c and d).

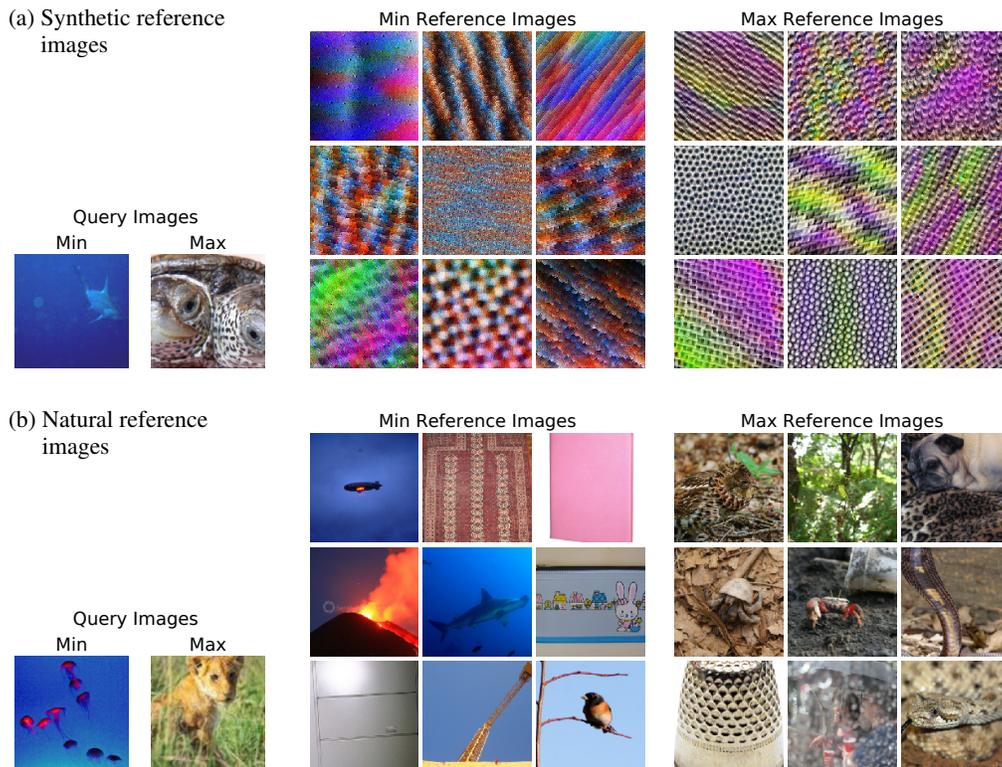


Figure 7: Another easy feature map (here: 3b, 1x1) from Experiment I where all subjects answered correctly for both synthetic and natural reference images. We are surprised that all the trials for synthetic reference images were answered correctly; to us, the decisive feature is not easily identifiable. The shown stimuli were shown to participant three.

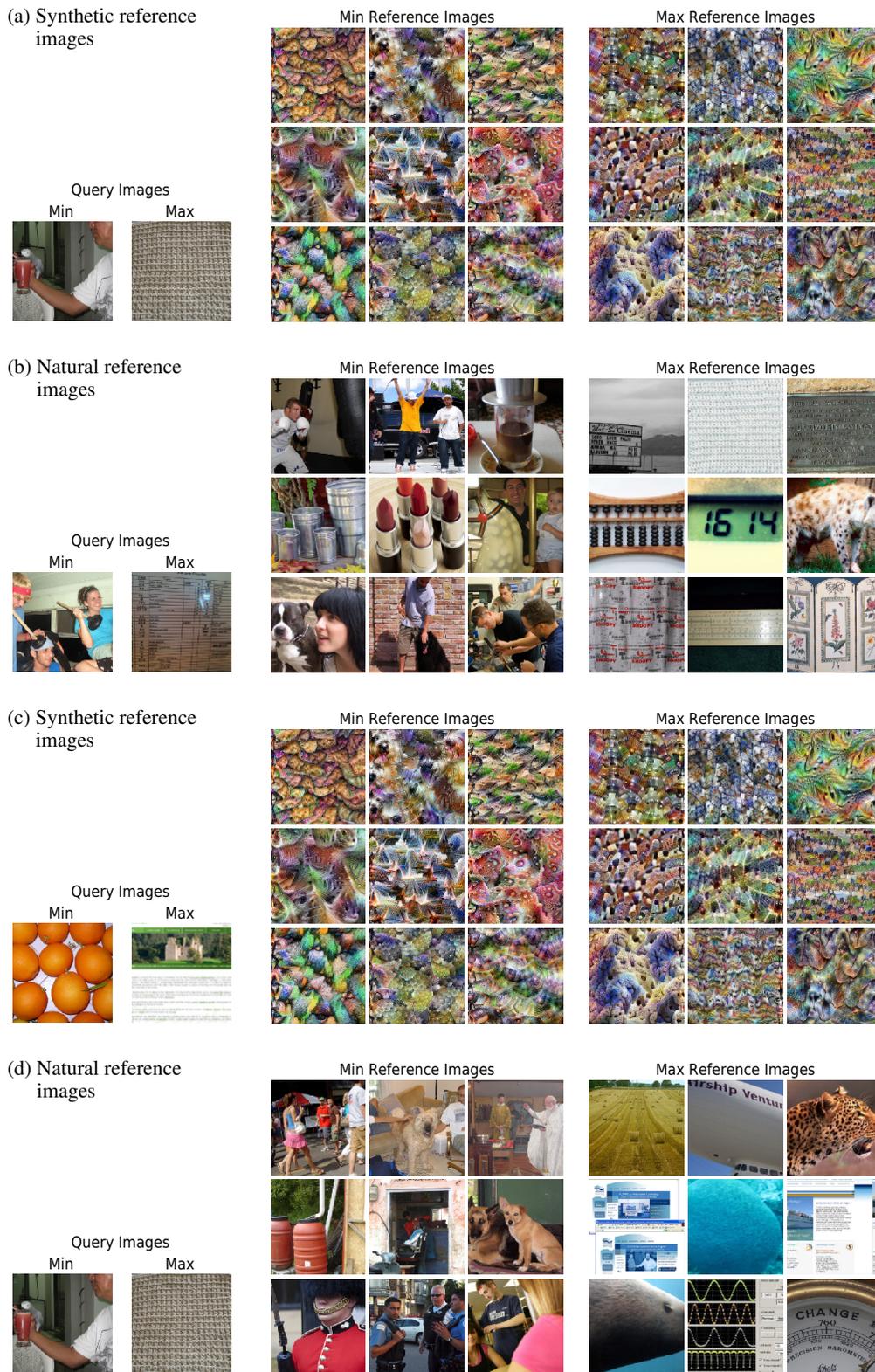


Figure 8: Example of a difficult feature map (4d, 5x5) from Experiment I where only four subjects answered correctly for synthetic reference images. The displayed stimuli were shown to participant two (a, b) and three (c, d), for stimuli shown to participant one, see Appendix Fig. 24.

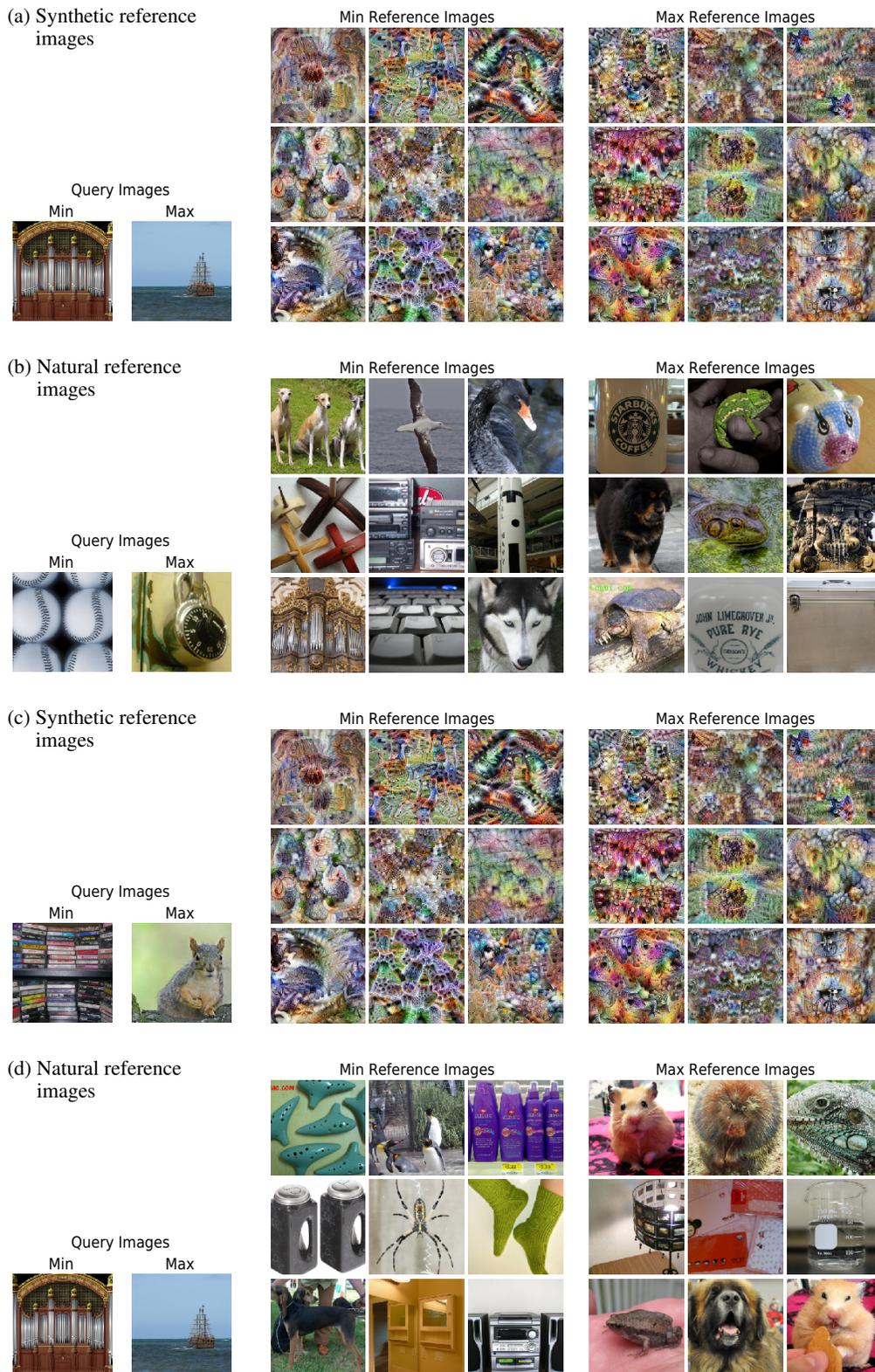


Figure 9: Example of a difficult feature map (5b, 5x5) from Experiment I where only four subjects answered correctly for natural (e-h) reference images. The displayed stimuli were shown to participant two (a, b) and three (c, d), for stimuli shown to participant one, see Appendix Fig. 24.

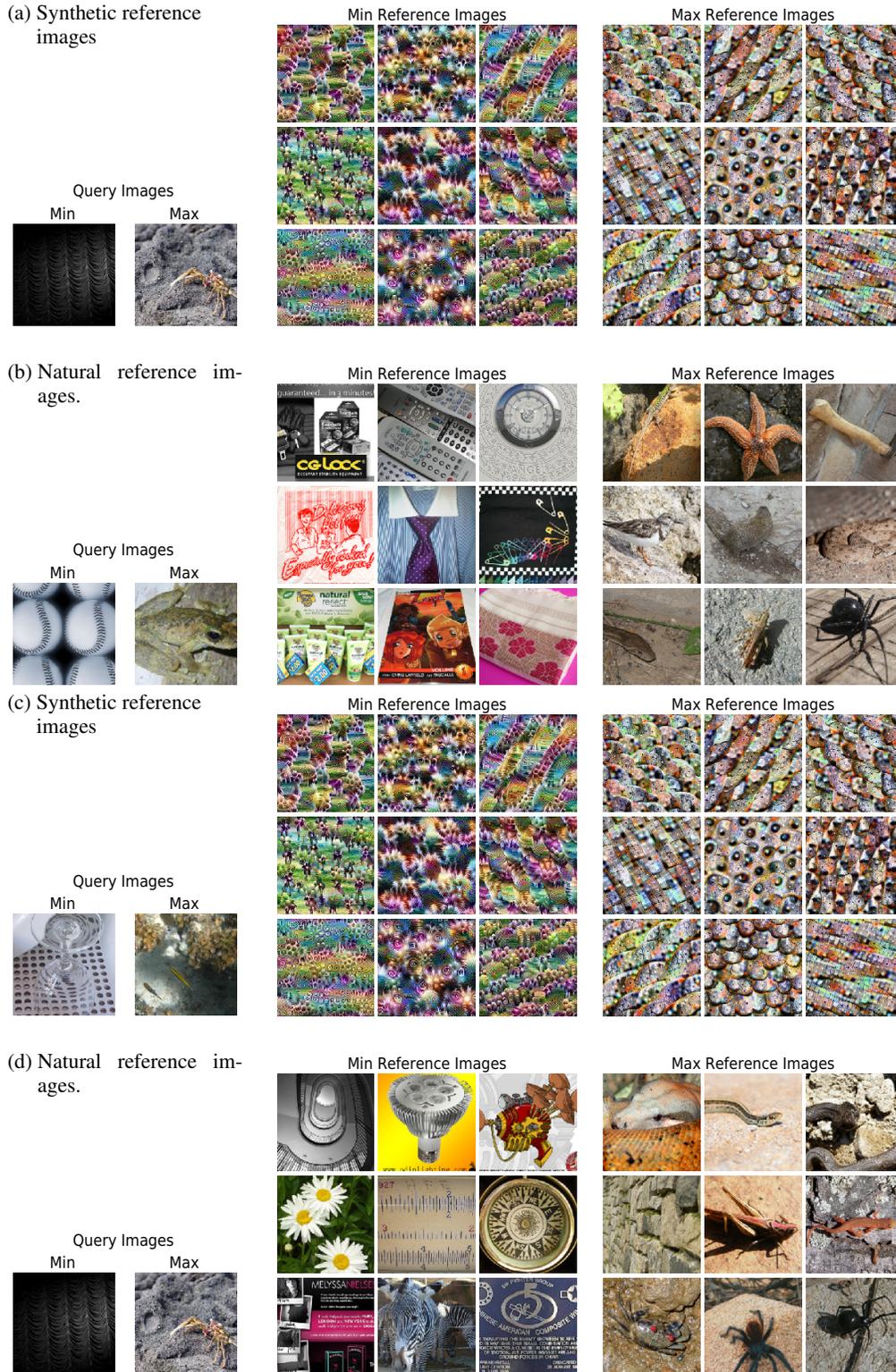


Figure 10: A feature map (here: 4a, Pool) from Experiment I where the feature is small (eyes) and a participant might perceive conflicting information (eyes and extremity-like structure in min reference images vs. eyes and earth-colors in max reference images). In this specific example, eight (nine) out of ten subjects gave the correct answer for this feature map given synthetic (natural) reference images. The displayed stimuli were shown to participants two (a and b) and three (c and d), for stimuli shown to participant one, see Appendix Fig. 25.

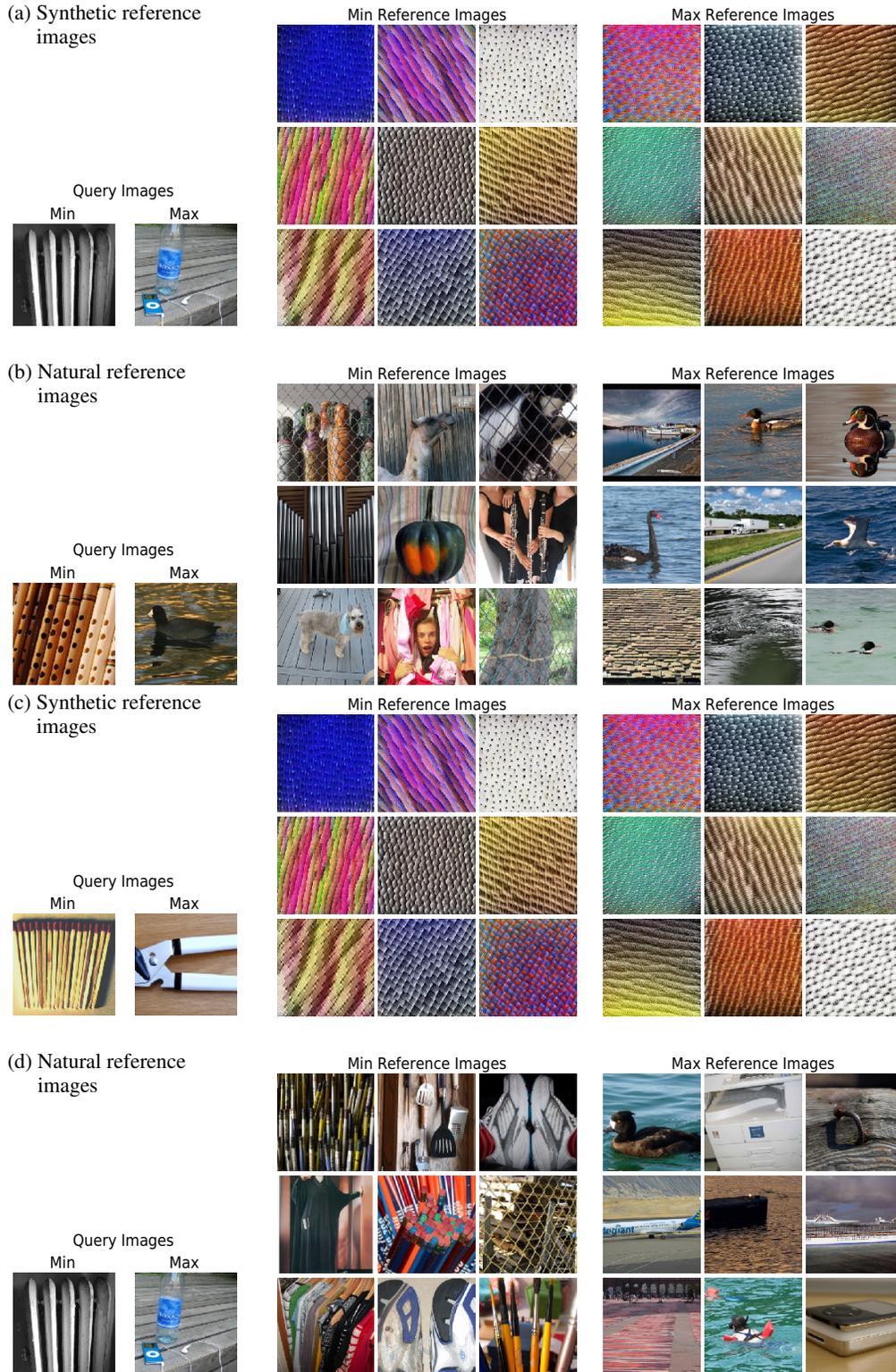


Figure 11: A feature map from a low layer (here: 3a, 3x3) from Experiment I where the feature seems to be a low level cue (horizontal vs. vertical striped) that is surprisingly clear in the natural, but surprisingly unclear in the synthetic reference images. In this specific example, seven (eight) out of ten subjects gave the correct answer for this feature map given synthetic (natural) reference images. The displayed stimuli were shown to participants two (a and b) and three (c and d), for stimuli shown to participant one, see Appendix Fig. 26.