Supplemental material
Bounded logit attention: Learning to explain image classifiers

A User study

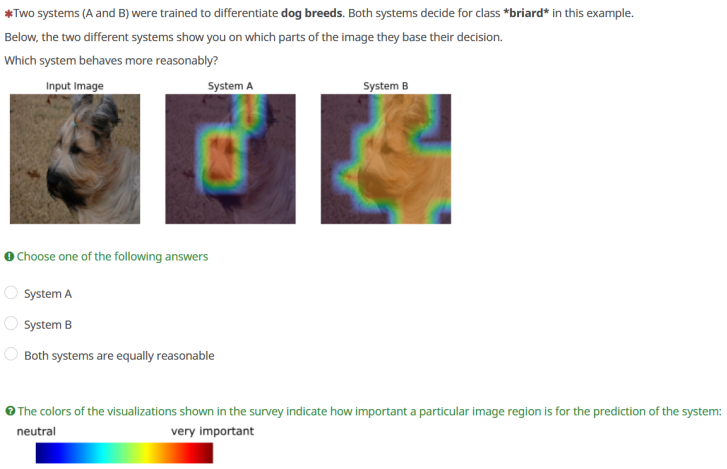

Figure S1: A question to users, here L2X-F ("System A", not to be confused with dataset A) vs. BLA-T ("System B"). Non-rearranged version of Figure 5.

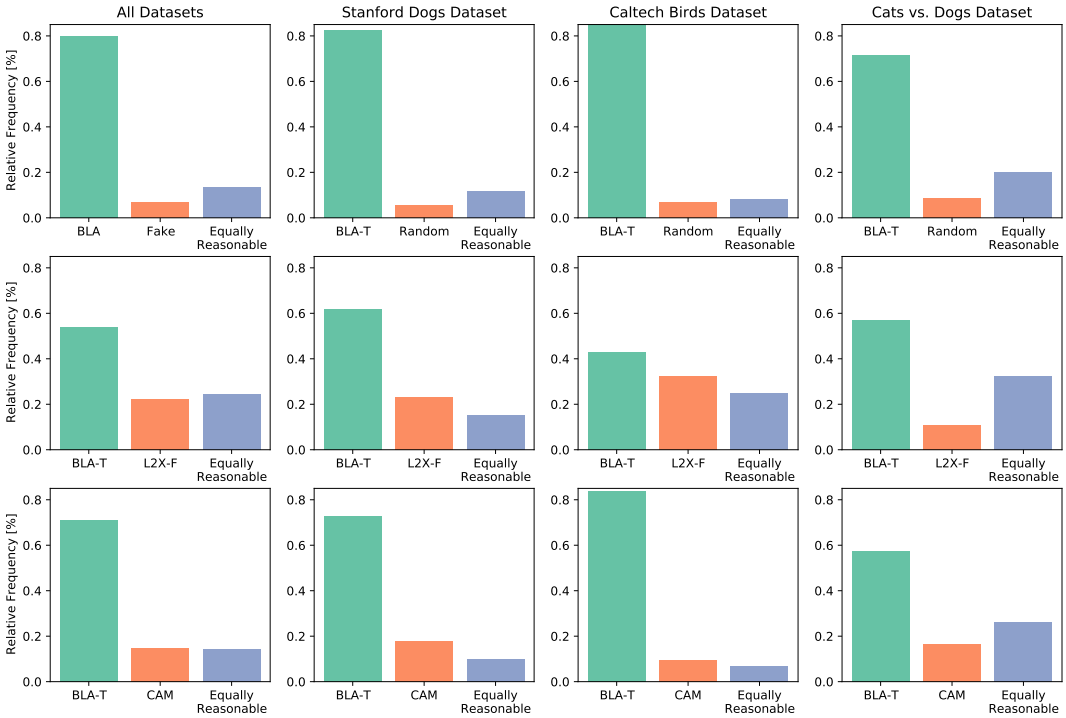

Figure S2: More detailed results of user study.

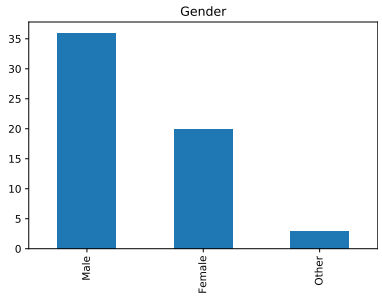

Figure S3: Gender of the study participants.

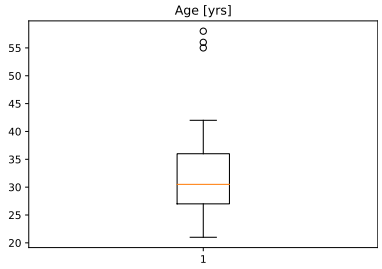

Figure S4: Age of the study participants.

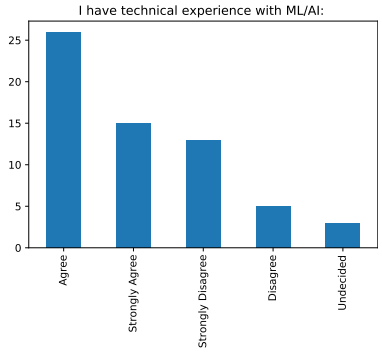

Figure S5: Machine learning/artificial intelligence (ML/AI) experience of study participants.

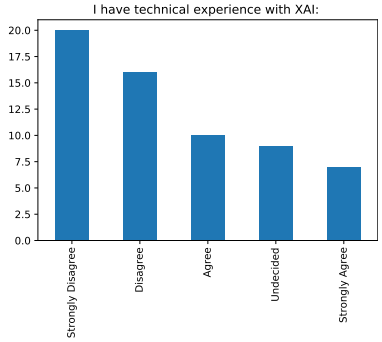

Figure S6: Explainable artificial intelligence (XAI) experience of study participants.

B Understanding the BLA explanation module

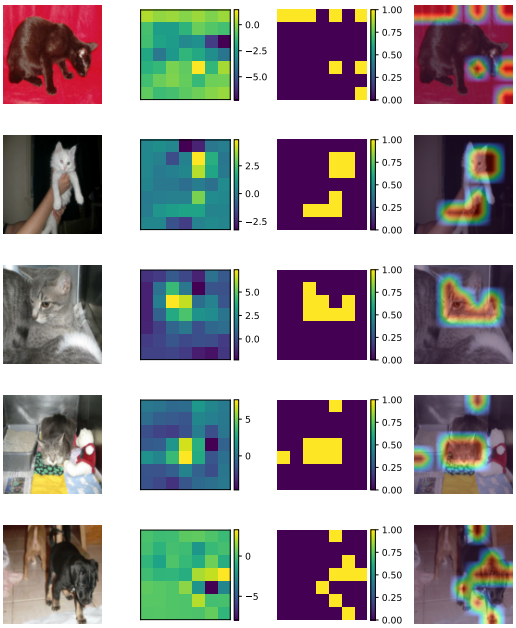

Figure S7: L2X-F – CvsD dataset

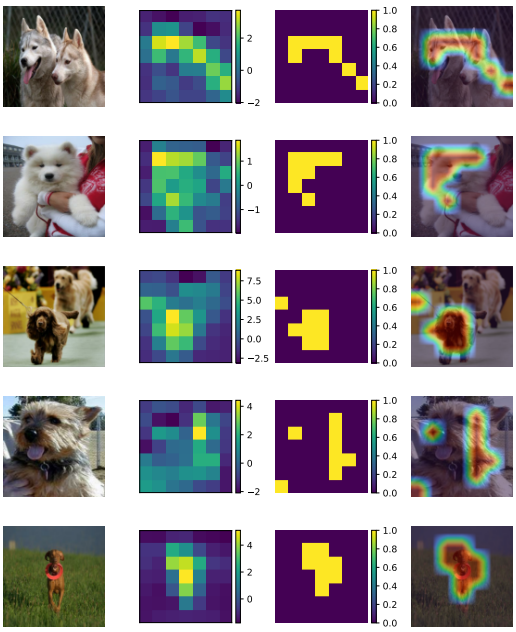

Figure S8: L2X-F – StanDogs dataset

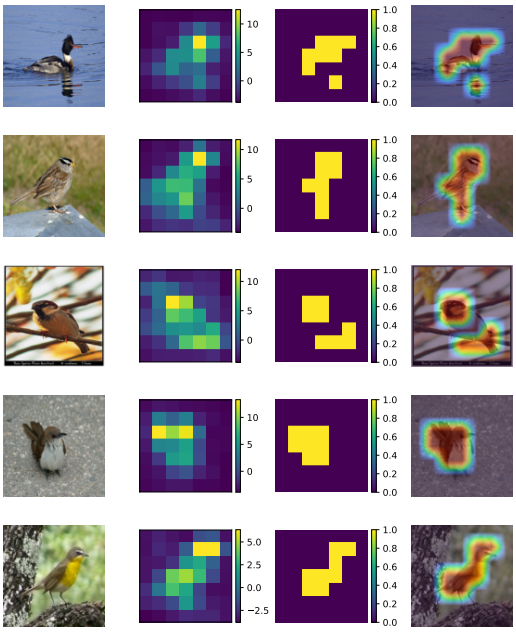

Figure S9: L2X-F – CUB-200 dataset

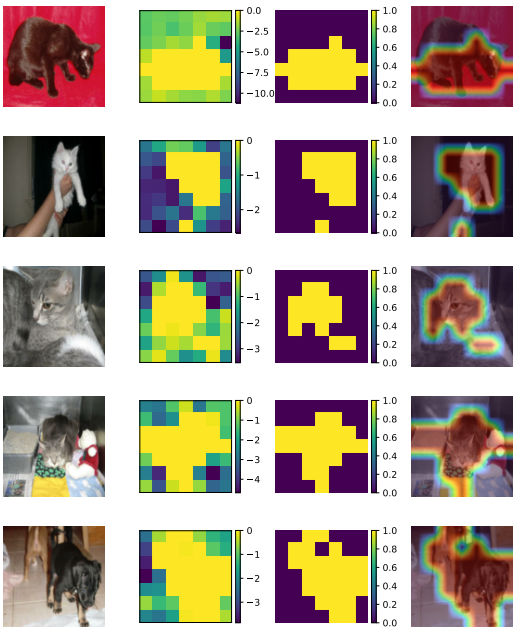

Figure S10: BLA – CvsD dataset

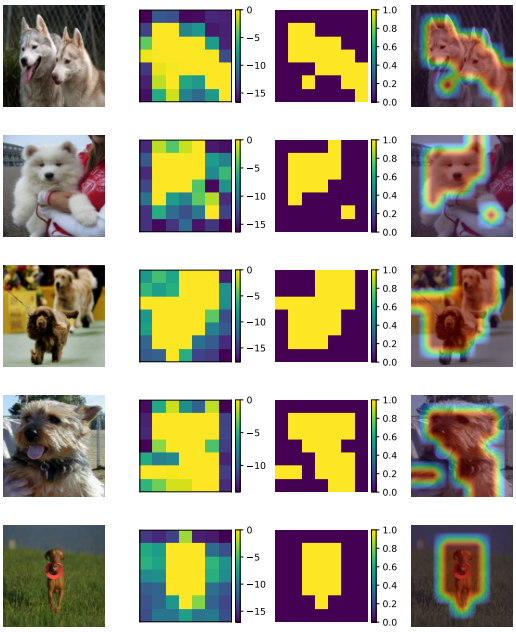

Figure S11: BLA – StanDogs dataset

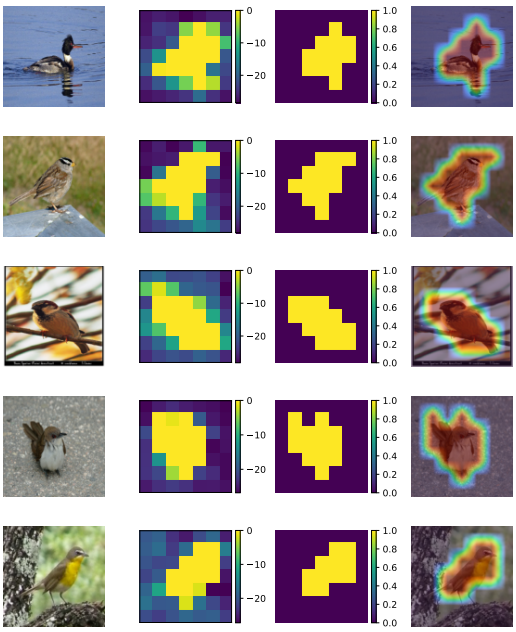

Figure S12: BLA – CUB-200 dataset

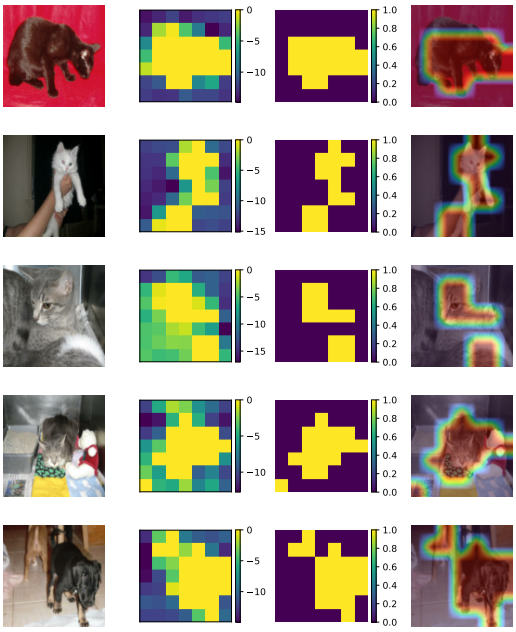

Figure S13: BLA-T – CvsD dataset

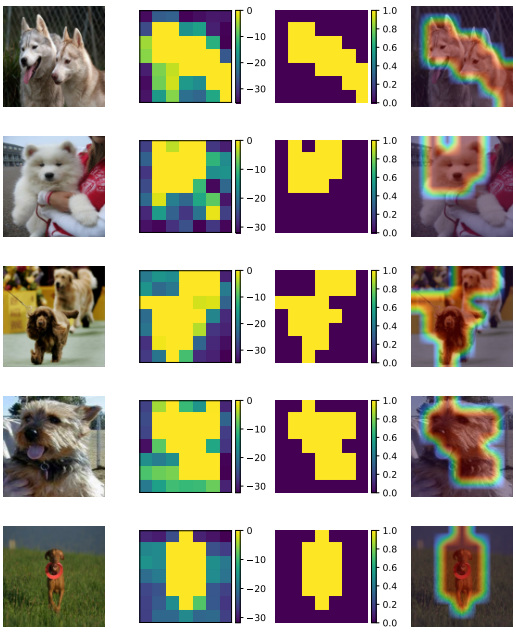

Figure S14: BLA-T – StanDogs dataset

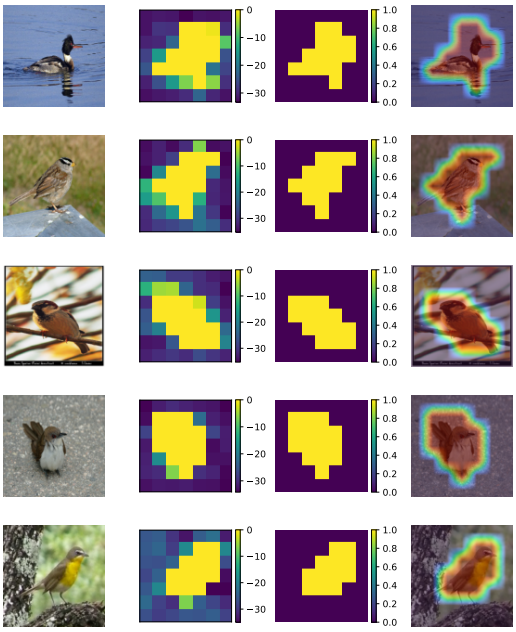

Figure S15: BLA-T – CUB-200 dataset

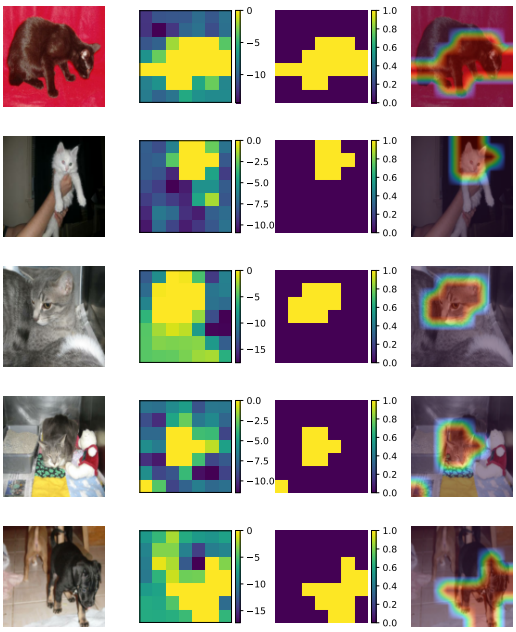

Figure S16: BLA-PH – CvsD dataset

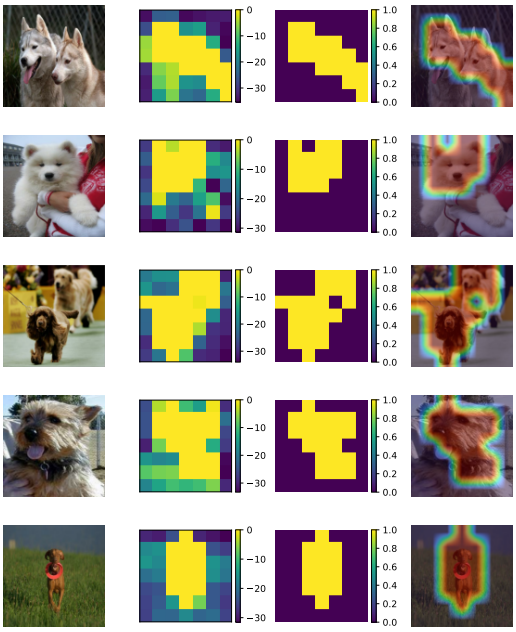

Figure S17: BLA-PH – StanDogs dataset

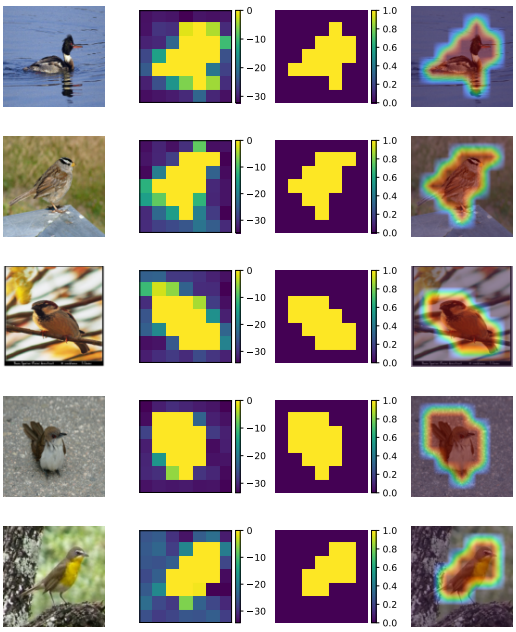

Figure S18: BLA-PH – CUB-200 dataset

C More context: attention with global concept vector

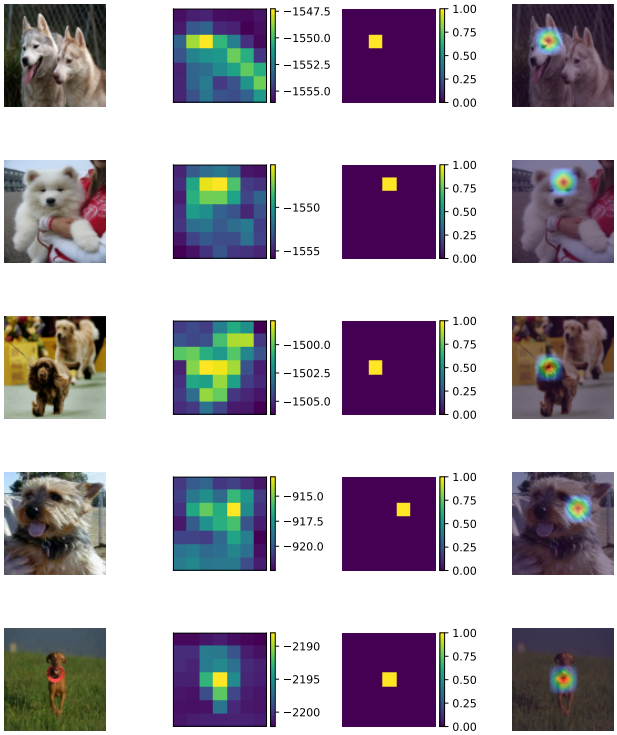

Figure S19: Attempt of using Jetley et al. (2018) for explainability.