We used Generalized Additive Models (GAM) to relate the number of images of a species and the recall of the model for that species using the mgcv package in R (Wood, 2018). The response variable was the base 10 logarithm of number of training images for a species and the explanatory variable was the recall (calculation of recall is described in Appendix S2) for that species. We used a Gaussian distribution for the errors and an identity link function. We used a spline smoother for the explanatory variable using GCV for unknown scale parameter, as is the default in the package mgcv. We chose the number of knots by using a penalty for each knot, allowing smoothing estimation to vary from 1 knot to 28 knots. We selected 28 as the maximum number of knots because this was the number of species or groups for which we evaluated recall (see Table 1). We then tuned the GAM by testing different smoothing scales, selecting the one that had the greatest support from the data using Akaike Information Criterion corrected for small sample size (AICC). This procedure resulted in a GAM that included 8 knots and fit the data relatively well, as = 0.529 and the model explained 56.5% of the deviance.

**References**

Wood, S. (2018). Package ‘mgcv’ (Version 1.8-24) [R].