Bayesian classification of affordances from RGB images
Lorenzo Mur-Labadía, Rubén Martínez-Cantin

Introduction
Affordances are the different action possibilities available in the environment depending on the motor and sensing capabilities of the individual [3]. They relate the objects, the actions and the possible effects of that actions carried on the objects [10]. Based on this, affordance prediction emerges as a powerful tool for autonomous and active agents where we need to understand the content of the scene: a cup is graspable, a road is traversable and a chair is sitable but it can also be graspable depending on the context.

Uncertainty estimation helps to discard low-confidence results, reasons about similarities, models noisy observations, analyses sources of uncertainty and serves as a basis for active learning algorithms.

Methods
Bayesian models predict the category and the degree of confidence of the prediction, providing a more robust tool for robotic applications[6, 11, 4, 1, 12]. We compare two alternatives:
• Monte-Carlo Dropout: approximates the posterior as the mean of the N forward passes during the test time with a random dropout of neurons, but we only train one single model
• Deep Ensembles: requires training M different models with random initialisation of their weights. Although we increase the training cost linearly, it works better when the posterior distribution does not follow a Bernoulli distribution.

The final prediction is the mean of the samples

\[
\hat{P}_m = \frac{1}{M} \sum_{m=1}^{M} P_m
\]

• Aleatoric uncertainty: it is associated with the noise inherent in the observations (motion noise, distant objects, boundaries) and it cannot be reduced by collecting more data [9].
• Epistemic uncertainty: related to the model knowledge, we reduce it by increasing the dataset [9].

Dataset
We conduct our experiments in the ADE20K dataset [2], composed of 44k objects, which was built on top of the ADE20K scenes [13], a popular semantic segmentation dataset. It divides the object-relationship interactions into 7 categories, including exceptions with social meaning:
1. Positive
2. Object non-functional
3. Physical obstacles
4. Socially awkward
5. Socially forbidden
6. The action is dangerous
7. Firmly negative

Metrics
We compute the mean accuracy of the predictions for the deterministic experiment

For the Bayesian experiments we report:
• Brier Score (BS): it measures the accuracy of the model. A perfect accurate model scores BS = 0, while a BS=1 means that the model is completely inaccurate
• Expected Calibration Error (ECE): it reports the calibration of the model, expressed as the difference between the confidence of the prediction and its accuracy.
• The evolution of the components of the covariance matrix: components in the trace reflect the variance of that category, while components out of the trace show inter-relationship between categories.

Conclusions
We propose a Bayesian deep learning model for affordance prediction directly from image data. We obtain higher performance over previous works and we extend the predictions with the quantification of the uncertainty at no cost in the classification. Comparing MC-Dropout and Deep Ensembles as the Bayesian techniques, we show an extensive analysis of the uncertainty estimation with the Brier Score, the ECE, the evolution of the components of the covariance matrix and a comparison of the epistemic and aleatoric uncertainty.

References