

1 We thank the reviewers for their detailed comments on our manuscript. We are glad that the response has been positive.
2 Below we respond to each reviewer individually, providing additional considerations (and data).

3 **Reviewer #1** recommends acceptance, highlighting the clarity of the paper, the interesting approach to classification,
4 and useful investigations of the cross-domain setting and kernel analysis. The reviewer pointed out two weaknesses,
5 stating that if they are addressed, this would increase the reviewer’s score. We thank the reviewer for the positive
6 feedback and suggestions for improving the work; we have addressed the requests as best as we could given the short
7 deadline. (1) *Experiments for out-of-range condition in ADKL, R2D2 and ALPaCA*. The code of ADKL is not available
8 and the code of R2D2 is only available for the classification case. For the in-range condition we have reported the
9 results from Tossou et al. (2019, reference in the paper). For ALPaCA there is public code for the in-range condition
10 (used to generate the MSE in Table 1). Following the request of the reviewer we have now modified the ALPaCA
11 code for out-of-range; the score of ALPaCA is 5.92 ± 0.11 (MSE, average of three runs). ALPaCA is better than
12 MAML (8.45 MSE) but significantly less accurate than our method (DKL has a lower MSE, 0.14 and 0.11). We
13 will add the new results to Table 1 and mark with an asterisk that the ADKL-R2D2 scores are from Tossou et al.
14 (2019). Porting the code of ALPaCA for the head-pose regression has been technically challenging given the short
15 deadline (it is based on completely different libraries). (2) *Quantitative evaluation of the uncertainty*. Following the
16 suggestion of the reviewer we ran experiments on model calibration. We have followed the protocol of Guo et al. (2017,
17 "On Calibration of Modern Neural Networks") which consists in temperature scaling plus estimation of the Expected
18 Calibration Error (ECE), a scalar summary statistic (the lower the better). For each method, we found the temperature
19 by minimizing NLL on logits/labels via LBFGS on 3000 tasks; then we estimated the ECE on the test set. We report the
20 results on the CUB dataset in the format "*1-shot/5-shot*" (percentage, average of three runs): DKL+BNCosSim (ours)
21 2.6/1.1, Feature Transfer 12.6/18.4, Baseline++ 4.9/2.0, MatchingNet 3.1/2.2, ProtoNet 1.1/0.9, MAML 1.1/2.5,
22 RelationNet 4.1/2.8. In 1-shot our model achieves one of the lowest ECE 2.6% beating most of the competitors (only
23 ProtoNet and MAML do better). In 5-shot our model achieves the second lowest ECE 1.1% (ProtoNet does marginally
24 better). These results provide additional evidence about the strength of the method in uncertainty estimation. We will
25 include tables in the appendix and a short summary in the paper.

26 **Reviewer #2** marked the paper for a clear acceptance, highlighting that the method is well motivated and performs well
27 in experimental comparisons. We thank the reviewer for the positive feedback which has captured the essential strength
28 of this work. Here we provide compact answers to the two questions. (1) *Potential overfitting*. For the standard conv-4
29 used in few-shot learning we did not encounter any overfitting problems. We have also tried a ResNet-10 (see Table 6,
30 supp. material) and similarly we did not observe overfitting. (2) *Use of regularization*. Following common practice in
31 few-shot literature, we have used data augmentation as the only form of regularization (for all methods). Note that, a
32 strength of our Bayesian approach is that it applies ML-II which has an implicit regularization effect.

33 **Reviewer #3** marked the paper as a good submission and recommended acceptance, highlighting as strengths the
34 efficiency of the method, the theoretical grounding, and the empirical evaluation. We thank the reviewer for the
35 rich commentary and considerations. (1) *Possibility of applying the method to Gaussian process-style models*. The
36 reviewer is right, the method could be easily extended to a standard Gaussian process by finding the parameters of the
37 kernel. (2) *On why the proposed approach works better than ALPaCA*. Our analysis suggests that ALPaCA may suffer
38 the shortcomings of complex meta-learning methods, being harder to train and less flexible in domain transfer and
39 out-of-distribution fitting. An additional experiment seems to confirm this hypothesis (see answer to Reviewer #1). In
40 the out-of-range condition ALPaCA has a higher MSE compared to our method (5.92 vs 0.11). Qualitative analyses
41 showed that ALPaCA is not able to provide a good fit for points out of its training range, with a behavior similar to
42 MAML (see Figure 1a in the paper). (3) *Other methods*. The reviewer asked for discussion of two uncertainty-aware
43 transfer learning methods. We thank the reviewer for indicating those papers, we will discuss them in the "Related Work"
44 section. (4) *Clarifications*. (i) "Small but related tasks", tasks sampled from a common distribution. (ii) "Dispersive",
45 colliding terminologies used in few-shot literature. (iii) "Needs to estimate higher-order derivatives", in MAML it is
46 necessary to estimate the gradient of the gradient in the training loop. (5) *Move Algorithm 1 to the main paper*. We
47 agree; this was our initial intention but given the space constraints we have been unable to do it gracefully. Given this
48 emphasis by the reviewer, we will try again for the camera ready.

49 **Reviewer #4** recommends acceptance, pointing out the efficiency and robustness of the method. At the same time the
50 reviewer identified the simplicity as one of its weaknesses. We discussed this point in the introduction of the paper,
51 stating how we consider simplicity as a strength. Previous methods are based on complex meta-learning routines, which
52 we showed can be greatly simplified by adopting our approach (with better performances). The few-shot learning
53 literature has been saturated with overly complicated solutions, in stark contrast we show that a well designed Bayesian
54 approach can prune unnecessary meta-learning loops while being very effective in practice. Simple is robust, easy
55 to implement, debug, reproduce, and apply. Though simple, this method is currently not common practice, and we
56 believe this paper strongly and rightly emphasizes the need to start with simple approaches before diving in to complex
57 meta-learning routines. We hope these considerations will resonate with the reviewer.