6 CONCLUSION

Through this thesis, we have sought to advance the integration between deep learning and probabilistic ML, with a focus on Gaussian processes (GPs) and Bayesian optimisation (BO). In this chapter, we reflect on our main contributions, discuss directions for future work, and conclude with a few parting words.

6.1 SUMMARY OF CONTRIBUTIONS

First, in Chapter 3, we improved upon prior work hyperspherical sparse GP approximation that uses nonlinear activations as inter-domain features, known as the ACTIVATED SVGP [65, 246], in which a single-layer feed forward NN emerges from the posterior predictive distribution. We provided an analysis of the limitations of this approach which *inter alia* preclude the use of widely-used covariance functions and nonlinear activations. Our key contribution was extending the orthogonally-decoupled sparse GP approximation to accomodate inter-domain features. We demonstrated that the combination of orthogonal inducing points and spherical activation features effectively mitigates the earlier limitations, not only bringing their predictive performance closer to NNs, but achieving superior scalability over alternatives.

Second, in Chapter 4, we provided an interpretation of CYCLEGANS as a Bayesian framework for inferring the hidden representations of entities from one domain as entities in another. Specifically, we framed the problem of learning cross-domain correspondences without paired data as inference in a LVM. First, we introduced the implicit LVM, where the prior over hidden representations is specified flexibly as an implicit distribution. We then introduced a new VI framework that differs from traditional VI in that it directly approximates the joint distribution based on a symmetrised KL divergence. Finally we showed that CYCLEGANS emerges as a variant of this framework, casting new light on this powerful class of deep generative models for image style transfer.

Third, in Chapter 5, we introduced a reformulation of BO based on solving a binary classification problem. We leveraged the connections between the improvement-based acquisition functions, density-ratio estimation (DRE), and class-probability estimation (CPE), to derive a binary classifier of candidate solutions that effectively serves as the acquisition function. By doing away with an explicit probabilistic model of the objective function, we eliminated the impediments posed by tractability requirements, enabling the seamless integration of deep learning and other powerful modelling paradigms in a manner that does not necessitate approximations or compromise scalability and representational capacity. Overall, this model-agnostic framework substantially expands the applicability of BO to diverse, challenging optimisation problem scenarios.

6.2 FUTURE DIRECTIONS

Looking to the future, promising new research avenues emerge that expand on the contributions made in this thesis.

ORTHOGONAL INTER-DOMAIN INDUCING FEATURES. Future work should explore additional combinations of inter-domain inducing features in the standard and orthogonal bases beyond the traditional inducing points and the spherical NN activation features examined here. In particular, since the orthogonal decoupling of GPs can be seen as a way to leverage different bases to separately represent the predictive mean and variance [217], it is promising to tap into the strengths of the spherical NN activation features [65, 246] and spherical harmonics features [64] to independently capture the predictive mean and variance, respectively. More broadly, to better enable exploration in this direction and accomodate the composition of various inter-domain inducing features, the software developed for this research should be refactored based on principles of modularity and separation of concerns. Adopting this more flexible and extensible design approach will facilitate seamless experimentation with combinations of diverse inter-domain inducing features in orthogonally-decoupled sparse GP approximations.

BORE BY DIRECT DRE. Future work should explore the potential benefits of other direct DRE methods. While the CPE approach is a big improvement from the problematic TPE approach, it is still thought to be a simple baseline in the DRE literature. The RULSIF [285] method may be of particular interest, not least because it is the only method of those discussed in Section 2.3 that directly estimates the *relative* density-ratio. Furthermore, since RULSIF is parameterised by a sum of Gaussian kernels, it enables the use of well-established mode-finding approaches, such as the *mean-shift* algorithm [44], for candidate suggestion. Finally, along the same avenue but in the opposite direction, one may also consider employing other DRE losses [166] for classifier learning, which would accomodate the use of powerful deep learning models.

EXTENDED BO BY CLASSIFICATION. Future work should explore extending BORE with model architectures suited for complex real-world optimisation problems. These “exotic” problems [70] include scenarios where the function involve *multiple outputs*, e. g., in multi-task, multi-fidelity, and multi-objective problems, where simple feed-forward NNs may be advantageous. They also include problems with *structured* or *sequential inputs*, where graph neural networks (GNNs) or Transformer architectures [268], respectively, may prove beneficial. By upgrading BORE to handle these complex optimisation problems with multiple outputs, structured inputs, and sequential inputs, it would become applicable to a wider range of challenging real-world tasks. Ultimately, the flexibility of the BORE means its capabilities continue to grow as researchers creatively integrate it with state-of-the-art modelling paradigms.

6.3 FINAL REFLECTION

Overall, this thesis has laid the necessary groundwork, improved existing frameworks, and offered new perspectives on Bayesian optimisation (BO), Gaussian processes (GPs), and deep learning. Our contributions advance the integration between deep learning and probabilistic ML, aiming to make decision support systems more capable, dependable, and equipped to handle the complex, dynamic, and uncertain challenges of the modern world.

In closing, we envision a future where the interplay between deep learning and probabilistic ML continues to evolve, leading to novel applications and breakthroughs that benefit society across a wide range of domains. The quest to realise the grand vision of AI – developing intelligent systems that can perceive, learn, decide, and act autonomously in complex real-world environments remains ongoing, and we are excited to contribute to this journey with our work. As the AI landscape continues to transform, the fusion between deep and probabilistic learning will undoubtedly play a pivotal role in shaping the AI of tomorrow.