

Chapter 1

Conclusion

Many aspects of the OceanQA dataset and the H-PERL approach have been discussed in previous chapters. This chapter therefore aims to summarise these details, and particularly focus on the strengths and weaknesses of our VideoQA architecture. We also attempt to compare our approach qualitatively to existing work in VideoQA. Finally, we conclude by outlining a number of possible extensions to our architecture, some of which may help to address the drawbacks of the approach, or apply the approach to different tasks.

1.1 Summary of H-PERL

In Chapter ?? we described the H-PERL architecture as a hybrid architecture - composed of both neural networks and symbolic reasoning modules - made of a pipeline of components. Each component in the architecture extracts a particular concept from the video or question. These concepts include: objects, including type and position; object properties, such as colour and rotation; binary relations between objects, such as close; and events between two consecutive frames of the video.

Chapters ?? and ?? each outlined specific H-PERL models. The hardcoded model, outlined in Chapter ??, used mostly manually engineered components. These components allowed the hardcoded model to find and correct errors in the object detection. The trained model, described in Chapter ??, continued to use some engineered components, but trained its *core* components using QA pairs. Chapter ?? outlined the performance differences between these models.

1.2 Comparison with Existing Models

1.3 Future Work