Eberhard Karls Universität Tübingen
Mathematisch-Naturwissenschaftliche Fakultät
Wilhelm-Schickard-Institut für Informatik

Bachelorarbeit Informatik

Florentin Doll

27. Oktober 2025

Gutachter

Prof. Dr. Martin Butz
Kognitive Modellierung
Wilhelm-Schickard-Institut für Informatik
Universität Tübingen

Prof. Dr. Nitram Ztub
Kognitive Modellierung
Wilhelm-Schickard-Institut für Informatik
Universität Tübingen

Betreuer

Tomáš Daniš
Kognitive Modellierung
Wilhelm-Schickard-Institut für Informatik
Universität Tübingen

Doll, Florentin:

Bachelorarbeit Informatik
Eberhard Karls Universität Tübingen
Bearbeitungszeitraum: 17.07.2025 – 17.11.2025

Abstract

Humans perceive the visual world through a process of selective attention, allowing them to focus high resolution vision on relevant parts of a scene and integrate that information across time. This thesis aims to model this process of visual attention with artificial agents, enabling them to efficiently process visual information in complex environments. I propose a novel (can i write novel here? it feels fancy but not sure) architecture that combines a recurrent visual attention mechanism with reinforcement learning to allow agents to learn where to look in order to maximize task performance. An auxiliary decoder helps to stabilize latent memory representations and enables interpretability of the current memory state. I evaluate the proposed architecture on a set of challenging visual tasks, including maze navigation, ...

Kurzfassung

Contents

1 Introduction **1**
2 Foundations **3**
3 Conclusion and Future Work **5**
A Important Additional Stuff **7**
Bibliography **9**

Chapter 1

Introduction

Human vision is inherently selective. Rather than processing the entire visual field uniformly, the human eye relies on a foveated view, where high-resolution vision is concentrated in a small central region (the fovea), while the surrounding peripheral vision is of lower resolution. Humans sequentially direct their gaze towards interesting or task-relevant regions of a scene, integrating information over time to form a coherent understanding of their environment. This mechanism of selective attention allows humans to efficiently process complex visual scenes under limited computational resources.

In contrast, most artificial vision systems process visual inputs in a uniform manner, requiring high computational resources to achieve comparable performance to human vision. They also usually only use a single feedforward pass to process an image, leading to limited interpretability and adaptability to changing environments.

Research in active vision and attention-based reinforcement learning has begun to address these limitations. Models such as the Recurrent Attention Model (RAM) (Mnih *et al.*, 2014) and its variants have demonstrated the potential of foveated vision and sequential attention mechanisms in artificial agents.

Chapter 2

Foundations

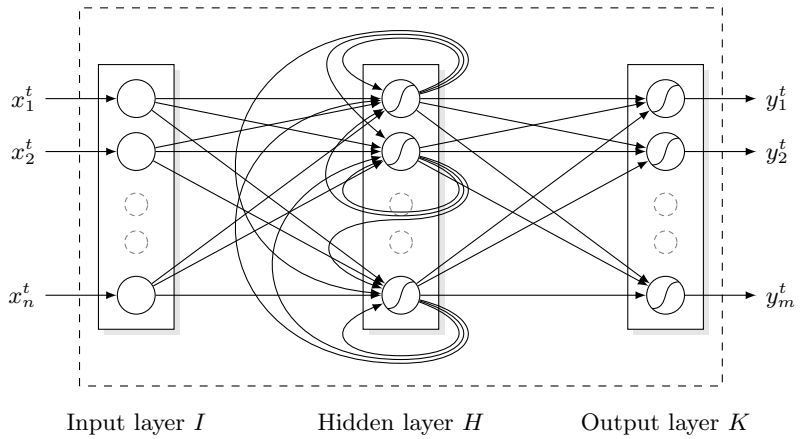

Figure 2.1: Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Table 2.1: Interesting table

Column A	Column B	Column C	Column D
1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Chapter 3

Conclusion and Future Work

Appendix A

Important Additional Stuff

Bibliography

Mnih, V., Heess, N., Graves, A., and Kavukcuoglu, K. (2014). Recurrent models of visual attention.

Selbständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe und dass alle Stellen, die dem Wortlaut oder dem Sinne nach anderen Werken entnommen sind, durch Angaben von Quellen als Entlehnung kenntlich gemacht worden sind. Diese Bachelorarbeit wurde in gleicher oder ähnlicher Form in keinem anderen Studiengang als Prüfungsleistung vorgelegt.

Ort, Datum	Unterschrift