

Overview

Building upon the motivation and goals established, this thesis is organized as follows:

We begin by introducing basic loss functions used in this work, for easy reference. These loss functions are essential for the concept of knowledge distillation, which we introduce next and is a key component of our approach. It follows an introduction to the Transformer architecture, which is the architecture used for all models in this work, followed by how Transformers can be used in multimodal models, and how We continue by exploring relevant approaches to multimodal models, including “. This chapter reviews the most influential models and methodologies, analyzing their strengths and identifying the limitations that hinder accessibility and scalability. This critical examination sets the foundation for our proposed approach.