Introspective Generative Modeling for Multimodal Translation
J.H.J. Linmans

Abstract

Motivated by the idea of exploiting the power of discriminative modeling for generative tasks, recent work on Introspective Neural Networks (INN) proposed a new framework to generate images using just discriminative models. In this thesis we introduce CWINN, the Conditional Wasserstein Introspective Neural Network used for multimodal translation tasks between text and images. In experiments on two benchmark datasets we show that the proposed model performs comparable with some of the existing state-of-the-art models for the task of text to image synthesis. Furthermore without redefining the network architecture, we show the application of CWINN to conditional image transformation: transforming an existing image based on a change in the corresponding caption. In order to apply CWINN to image captioning tasks, we present a continuous representation of natural language based on latent factors and second we introduce Introspective Language Models (ILM) used to model this representation. In a proof of concept study on a dataset of captions, we demonstrate compelling results using both word and character level ILMs to generate sentences. Lastly we analyze the potential of CWINN for image captioning by adopting ILMs.