

## Master thesis : Energy-based Multi-Modal Attention

**Auteur :** Werenne, Aurélien

**Promoteur(s) :** Maree, Raphael

**Faculté :** Faculté des Sciences appliquées

**Diplôme :** Master en ingénieur civil en informatique, à finalité spécialisée en "intelligent systems"

**Année académique :** 2018-2019

**URI/URL :** <http://hdl.handle.net/2268.2/7854>

---

### Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

---

# Master Thesis: Energy-base Multi-Modal Attention

Name: Aurélien Werenne

Supervisor: Dr. Raphaël Marée

Section: Computer Engineering - Intelligent Systems

Academic Year: 2018-2019

## Summary

A multi-modal neural network exploits information from different channels and in different terms (e.g., images, text, sounds, sensor measures) in the hope that the information carried by each mode is complementary, in order to improve the predictions the neural network. Nevertheless, in realistic situations, varying levels of perturbations can occur on the data of the modes, which may decrease the quality of the inference process. An additional difficulty is that these perturbations vary between the modes and on a per-sample basis. This work presents a solution to this problem. The three main contributions are described below.

First, a novel attention module is designed, analysed and implemented. This attention module is constructed to help multi-modal networks handle modes with perturbations.

Secondly, two new regularizers are developed to improve the generalization of the robustness gain on more intensive failing modes (relative to the training set).

Lastly, a unified multi-modal attention module is presented, combining the main types of attention mechanisms in the deep learning literature with our module. We suggest that this unified module could be coupled with a prediction model to enable the latter face unexpected situations, and improve the extraction of the relevant information in the data.