Multimodal-based adversarial defence

Background: Adversarial attacks are small bounded-norm perturbations of a network's input that aim to alter the network's output and are known to mislead and undermine the performance of deep neural networks (DNNs). Adversarial defenses then aim to mitigate the effect of such attacks.
Relevant papers: “Explaining and harnessing adversarial examples”, “Can audio-visual integration strengthen robustness under multimodal attacks”.
Project Description: In this project, we discuss adversarial attacks on Multimodal models and aim to utilize the various input channels for improved adversarial robustness.
Prerequisites: Deep learning course
Supervisor: Yaniv Nemcovsky (yanemcovsky@gmail.com)