Prof. R. Wattenhofer

Teaching Blind Vision Models to See

Language and vision models have shown themselves to be excellent at many tasks over the past few years. However, they are still nowhere near general intelligence which, for examples, has been shown in a recent paper.

Here, the authors asked state-of-the-art models to solve simple tasks, you can see some examples of the tasks in the figure (counting number of line crossings or number of squares).

This project will focus on teaching vision models to solve these tasks, and evaluating their performance in more general settings.

The goal of this is to determine if these tasks are viable as CAPTCHA challenges, online tests of whether a user is human or not. The problem here is that there should be enough variability in the input space that a model would struggle.
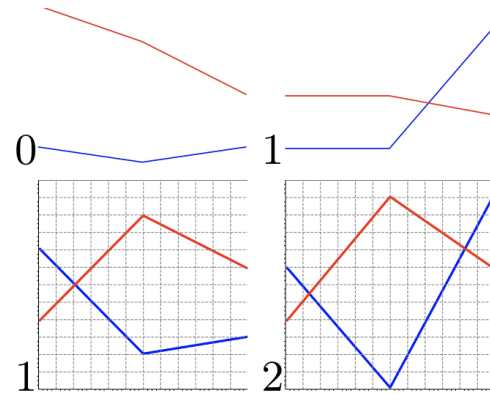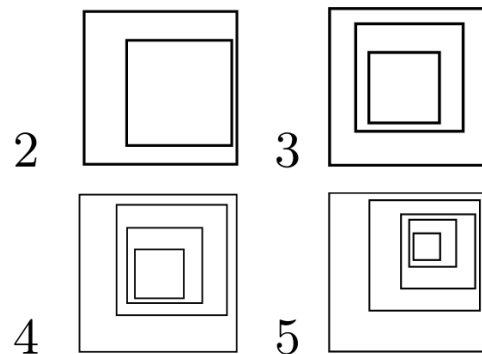

Figure 1: Count line crossings

Figure 2: Count nested squares

Requirements

Programming skills (Python, C / C++, etc.) and a good knowledge of machine learning and machine learning libraries.
We will have weekly meetings to address questions together, discuss progress, and think about future ideas.

Contact

In a few short sentences, please explain why you are interested in the project and about your coding and machine learning background (i.e., your own projects or relevant courses you have taken at ETH or elsewhere).

- Andreas Plesner: aplesner@ethz.ch, ETZ G95