Introduction

In this section of the course you will be writing a basic convolutional neural network (CNN) for handwritten digit recognition of the MNIST dataset. We will take an in-depth look at a model based off of Yann LeCun's LeNet-5.

A. Digit recognition

When we start building an image recognition model for the first time, it's usually a good idea to train and evaluate it on a relatively simple dataset. This allows for less complicated models, quicker training, and less image preprocessing. Basically, we want to learn the inner workings of a convolutional neural network before we even try to apply it to more complex tasks.

One of the simplest tasks we can perform is handwritten digit recognition. Given an image of a handwritten digit (i.e., 0, 1, ..., 9), we want our model to be able to correctly classify its numeric value. Though this task seems relatively simple, it is actually used fairly often in real life, such as automatically extracting credit card numbers from a picture. The dataset we will use for digit recognition is the MNIST dataset, which is the de facto dataset used for machine learning-based digit recognition.

B. CNN dominance

Nowadays, every task in the field of image recognition is dominated by convolutional neural networks (CNNs). Since the introduction of LeNet-5, the pioneering CNN, in 1998 for digit recognition, CNN architectures have become king in nearly all computer vision tasks. The CNN forms the backbone of many of today's cutting-edge technologies, such as self-driving cars, robotics, and facial recognition.

Although CNNs dominate image related tasks, we could still theoretically use a multilayer perceptron model for classifying images. However, the MLP would not perform nearly as well as a CNN. So what exactly makes a CNN special for

image input data? The key to the model's success is its *convolutional layers*, which is one of the main topics we'll discuss in the following chapters.