

In-Class Activity

Lecture 3 & Lab 2 Review: Breaking CAPTCHAs with CNNs

Questions

1. CAPTCHA (CNN Approach + Dimensions)

Figure 1 shows a CAPTCHA prompt: “*Select all images with a school bus.*” In this activity, your goal is to solve **only this type of CAPTCHA** (school bus vs. not school bus).

A CAPTCHA presents a **30×30 RGB** image split into a **3×3** grid of **non-overlapping 10×10 tiles** (9 tiles total). You want to design a CNN-based approach to solve this CAPTCHA.

- (a) State the **number of classes** your model should predict for the school-bus task.
- (b) Briefly list the key **preprocessing** steps before inference.
- (c) Decide what you will feed into the CNN. State the **CNN input dimension** for your approach.
- (d) For your approach in (c), state the **CNN output dimension** (include shape).
- (e) Briefly describe the **postprocessing** procedure to convert model outputs into a final CAPTCHA solution.

2. Convolution and Average Pooling

Consider the 5×5 feature map X and 3×3 filter K :

$$X = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \quad K = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Use **valid convolution** (no padding) with **stride = 1**. Let $Y = X * K$.

- (a) State the **dimension** of Y .
- (b) Compute the **entire** output feature map Y .

Now apply **average pooling** to Y using a **2×2 window** with **stride = 1**. Let the pooled output be P .

- (c) State the **dimension** of P .
- (d) Compute the **entire** pooled output P .

3. Softmax

Given logits $\mathbf{z} = [2, 0, -1]$:

- (a) Compute $\text{softmax}(\mathbf{z})$.
- (b) State the predicted class (index 1, 2, or 3).



Figure 1: CAPTCHA example used in this question.