

Hipótesis para la Solución del Reconocimiento de Captchas

Grupo 5

Hector Ivan Tovar Jaimes

Javier Pinto Losada

Juan Jose Restrepo

1. Planteamiento del Problema

El reconocimiento de texto en imágenes tipo CAPTCHA constituye un problema complejo en el ámbito de la visión por computador, dado que los caracteres aparecen distorsionados, embebidos en ruido y con espaciado irregular.

2. Solución del Problema

A través de lo aprendido en esta unidad, proponemos una solución basada en redes neuronales convolucionales (CNN), específicamente diseñadas para tareas multiclase y multisalida, las cuales han demostrado ser eficaces en clasificación de imágenes.

3. Requerimientos de los datos

Para entrenar un modelo efectivo, se requiere una base de datos con las siguientes características:

- Imágenes de CAPTCHA con distorsiones realistas, variaciones en el fondo, ruido, inclinación y curvatura de texto.
- Etiquetas asociadas a cada imagen que contengan la secuencia completa de caracteres (por ejemplo, "T8v5E").
- Resolución estándar.
- Variabilidad en los caracteres y combinaciones para evitar sobreajuste.
- Balance en la aparición de letras y dígitos del 0 al 9.

4. Tipo de Red y Configuración

Opción Elegida: CNN (Red Neuronal Convolucional)

La arquitectura CNN es adecuada por su simplicidad y efectividad en la extracción de características espaciales, utilizando bloques de capas convolucionales con la siguiente arquitectura:

- Kernel: (3x3)
- Activaciones: ReLU
- Padding: same
- 16,32 y 32 neuronas para cada una de las capas convolucionales
- (03) capas de Max Pooling 2D
- (01) Capa de Batch Normalization
- (01) Capa Flatten
- Arquitectura convolucional es de Tipo Funcional.

Estas características la convierten en una opción sólida para el reconocimiento de patrones complejos como los presentes en los CAPTCHAs.

5. Entradas y salidas del modelo.

Entrada: Imagen con dimensiones especificadas por `input_shape`. (50,200,1) donde el 50 es el alto, 200 es el ancho y 1 representa que son imágenes a blanco y negro.

Salida: Son 5 capas de 36 neuronas cada una, con activación `softmax`, correspondiente a 36 clases.

Transformación de Etiquetas:

Para estructurar el problema como una clasificación multisalida, se propone una transformación de las etiquetas que permita al modelo predecir múltiples caracteres simultáneamente. Este enfoque incluye los siguientes pasos:

1. Transposición de etiquetas para reorganizar los datos y obtener columnas, donde cada una representa un carácter del CAPTCHA.
2. Expansión de dimensiones para adaptarse a los requisitos del framework:
`columnas1 = np.expand_dims(np.array(columnas[0]), -1)`, etc.
3. Concatenación para formar una matriz de salidas:
`Y = np.concatenate((columnas1, columnas2, columnas3, columnas4, columnas5), axis=1)`

Este enfoque permite entrenar un modelo de red neuronal con múltiples salidas simultáneas, mejorando la eficiencia y precisión en la clasificación de cada carácter del CAPTCHA.

Preprocesamiento

- Conversión de la imagen a escala de grises.
- Redimensionamiento al tamaño definido.
- Normalización de los valores de los píxeles a $[0, 1]$.
- (Opcional) Aumento de datos: rotaciones, ruidos, escalado, desplazamientos, etc.

Métricas de Evaluación

- Exactitud por carácter: porcentaje de caracteres individuales correctamente clasificados.
- Exactitud completa: porcentaje de CAPTCHAs completamente correctos (todos los caracteres en la posición exacta).

4. Conclusión

La solución planteada, basada en una red convolucional adaptada para clasificación multisalida, permite enfrentar eficazmente el desafío del reconocimiento de texto en imágenes CAPTCHA. El uso de la transposición de etiquetas y la estructuración de múltiples salidas independientes mejora la eficiencia del entrenamiento y la precisión del modelo. Esta propuesta aprovecha de forma coherente todos los elementos vistos en la unidad: capas convolucionales, pooling, arquitectura CNN y principios de clasificación multiclase. Por lo anterior, se concluye que la solución planteada es compatible con los datos disponibles.