Week 8 Research Stay

Contexto

En esta presentación se comparan los papers proporcionados, que tratan de los modelos encoder-decoder. Las comparaciones se centrarán en las referencias 3-4.

Metodología de búsqueda

Texto encontrado en Google Scholar. Buscado más específicamente relacionado con HTR

Week 8 Research Stay

Términos Preliminares

- **sequence-to-sequence (seq2seq**): machine learning models used for tasks like machine translation, speech recognition, and text summarization.
- **CNN (convolutional neural network**): arquitectura de red para el aprendizaje profundo que aprende directamente de los datos. Usa aprendizaje supervisado que procesa sus capas imitando al cortex visual del ojo humano para identificar distintas características en las entradas que en definitiva hacen que pueda identificar objetos y “ver”.
- **Attention matrix**: Componente de los mecanismos de atención de las NNs que determina la importancia de cada elemento de una secuencia en relación con otros, lo que permite que el modelo se centre en las partes relevantes de la entrada al generar salidas.
- **Sequence learning tasks**: tasks like machine translation, image captioning and speech recognition.
- **HTR**: Handwritten text recognition.
- **Thin plate splines (TPS):** técnica basada en spline para la interpolación y suavizado de datos. (Un spline es una curva diferenciable definida en porciones mediante polinomios)
- **Puntos fiduciales (fiducial points):** Usado como punto de referencia.
- **ResNet (Red Residual Neuronal):** modelo de aprendizaje profundo en el que las capas de pesos aprenden funciones residuales con referencia a las entradas de las capas.

Week 8 Research Stay

Comparación [1]

"Evaluating Sequence-to-Sequence Models for Handwritten Text Recognition" (Ref. 2) Se centra en demostrar qué tan eficientes son diferentes tipos de modelos Seq2Seq para e; HTR (Handwritten Text Recognition), mientras que el artículo que encontré, "AttentionHTR: Handwritten Text Recognition Based on Attention Encoder-Decoder Networks" (Ref. 4)(un poco más reciente) busca presentar un solo modelo diseñado para optimizar el HTR.
Otra diferencia destacable entre estos dos documentos es su objetivo. Ref. 3 se centra más en comparar los resultados entre diferentes modelos, mientras que Ref. 4 busca específicamente mejorar la precisión y eficacia de los modelos basados en atención.

Ambos usan Attention-based encoder-decoder network con una estructura similar; encoder-attention m.-decoder, pero la principal diferencia entre estos dos documentos son las "etapas" o "partes" en las que se desglosan estas arquitecturas que se usaron para HTR.

Week 8 Research Stay

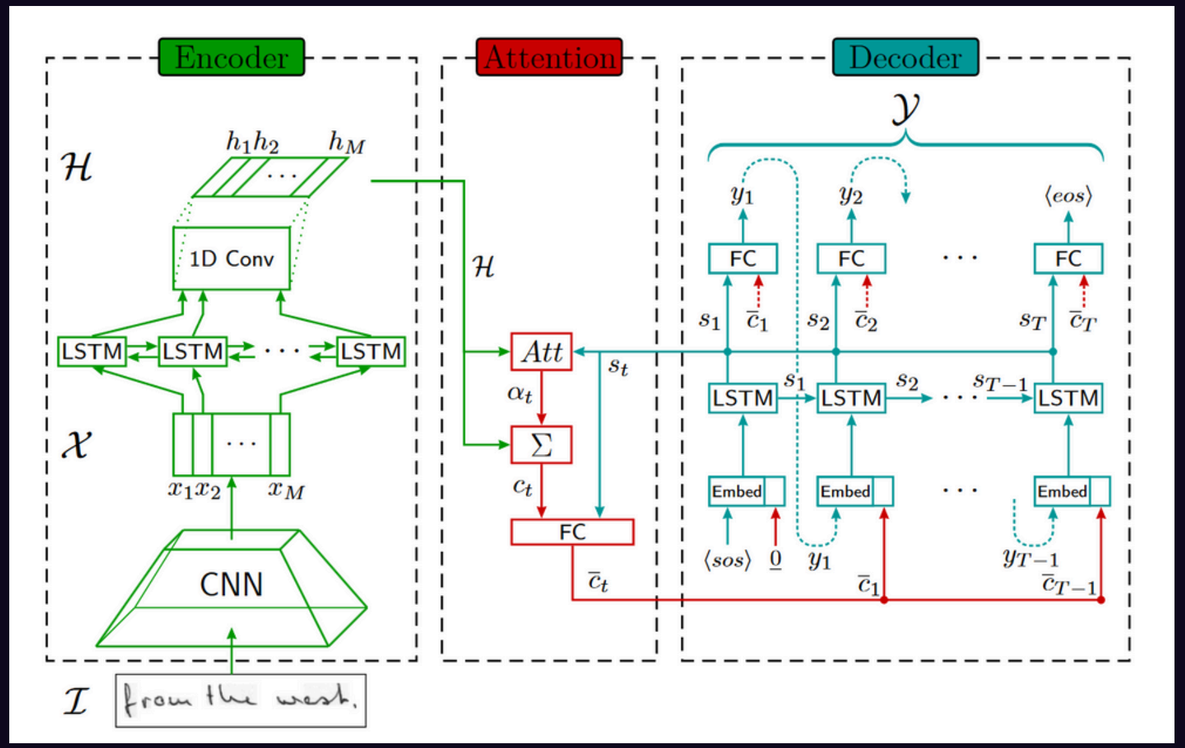
Comparación [2]

Modelo propuesta por Ref. 3:

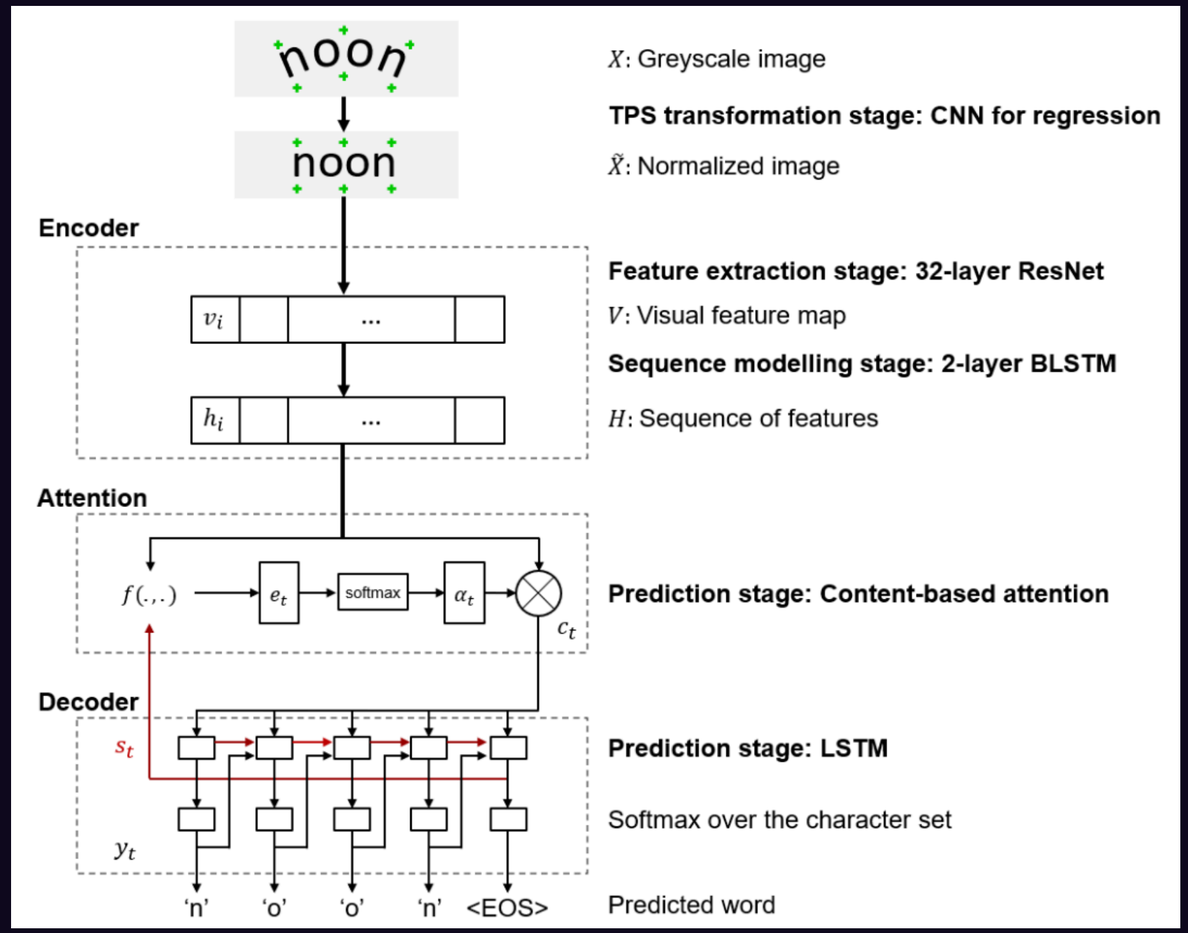
1. Codificador que combina una CNN como un extractor de características genéricas con capas recurrentes para introducir un contexto temporal en la representación de características
2. Decodificador que utiliza una capa recurrente para interpretar esas características 3. Mecanismo de atención que permite al decodificador centrarse en las características codificadas más relevantes en cada paso de tiempo de decodificación.

Modelo propuesta por Ref. 4:

1. transformación: Las imágenes de palabras de entrada se normalizan mediante TSP (Thin plate splines), esto toma las coordenadas de los puntos fiduciales que se utilizan para capturar la forma del texto. Las coordenadas de los puntos fiduciales hacen regresión mediante una CNN y la cantidad de puntos es un hiperparámetro.
2. Extracción de características: Se utiliza una red neuronal residual de 32 capas (ResNet) para codificar una imagen de entrada en escala de grises normalizada de 100 × 32 en un mapa de características visuales 2D.
3. Modelado de secuencias: Las características V de la etapa de extracción de características se transforman en secuencias de características H, donde cada columna en un mapa de características
4. Predicción: Se utiliza un decodificador basado en la atención para mejorar las predicciones de secuencias de caracteres. El decodificador es un LSTM unidireccional y la atención se basa en el contenido.



Arquitectura de Ref. 3



Arquitectura de Ref. 4

Week 8 Research Stay

Referencias

- [1] Sentence-Level Grammatical Error Identification as Sequence-to-Sequence Correction
- [2] Understanding How Encoder-Decoder Architectures Attend
- [3] Evaluating Sequence-to-Sequence Models for Handwritten Text Recognition

- [4] AttentionHTR: Handwritten Text Recognition Based on Attention Encoder-Decoder Networks