Identificación de figuras musicales 2D utilizando métodos de Deep Learning, una comparación entre Faster R-CNN y Mask R-CNN

Juan Camilo Montañez Osorio, Arnold Julian Herrera Quiñones, Roger Enrique Guzmán Avendaño

**Abstract**

**(pendiente)**

1. **Planteamiento del problema**

La música es un área que ha sido inherente al ser humano a lo largo de los años, y que en la actualidad la mayor parte de la población toma como una distracción o pasatiempo, mientras que otra parte de la población la toma como un estilo de vida para su sustento económico o de manera creativa y/o artístico de la persona. El aprendizaje de la música se realiza mediante diferentes maneras entre las que se encuentran el aprendizaje autodidacta y las clases magistrales en escuelas.

La representación escrita de la música se hace mediante las partituras, cuya finalidad es representar el ritmo, el tiempo y la tonalidad que se debe seguir dentro de una canción. Las partituras se componen de un pentagrama y sobre él se ubican las diferentes figuras musicales, la lectura de las partituras tiende a ser difícil de leer de acuerdo con la complejidad de las notas que se pretendan representar que en algunos casos genera la fusión de dos o más figuras musicales dificultando el correcto reconocimiento de estas.

Actualmente existen diversas herramientas tales como NoteReader, SheetVision y MusicPal, que, aunque sirven de ayuda para la identificación reconocimiento de las figuras, sin embargo, tienen ciertas falencias en la segmentación de las figuras musicales que tienen tamaños reducidos o que se encuentran en imágenes que no tengan buena resolución. Por lo que se hace necesario buscar unos métodos de segmentación que sean más precisos y comparándolos con el método propuesto en (Osorio, 2019), para lograr determinar que método de identificación de figuras musicales tiene un mejor rendimiento y exactitud.

1. **Trabajos Relacionados**

* Método de identificación automática para reconocimiento de elementos musicales presentes en imágenes 2d de partituras usando técnicas de aprendizaje profundo**:**

En este proyecto se propone una estrategia metodológica que implementa una arquitectura Faster -RCNN enfocado a la identificación y detección de objetos según la figura musical presente en el pentagrama. (Osorio, 2019)

* **TOWARDS FULL-PIPELINE HANDWRITTEN OMR WITH MUSICAL**

**SYMBOL DETECTION BY U-NETS**

En este articulo se menciona diversas técnicas para el reconocimiento de las figuras musicales tales como Convex Hull Segmentation que ayuda con la detección de objetos que pueden contener espacios. Así mismo utilizan las redes U-nets que complementan las redes de contratación que tienen capas sucesivas

* **MNR: MUSCIMA Notes Recognition. Using Faster R-CNN on handwritten music dataset**

La detección precisa de símbolos musicales en imágenes de partituras impresas orquestales o de piano históricas, complejas y densas puede ser un desafío debido a las antiguas técnicas de impresión o la degradación del tiempo. Debido a que los problemas de segmentación pueden variar ampliamente, se necesita un enfoque basado en datos como el uso de detectores de aprendizaje profundo. Sin embargo, la producción de anotaciones de detección (cuadros delimitadores de símbolos + clases) para dichos sistemas es costosa y lleva mucho tiempo. Proponemos entrenar dicho modelo con datos sintéticos y anotaciones producidas por un programa de composición tipográfica de música. Analizamos qué clases son relevantes para la tarea de detección y presentamos una primera selección de archivos de composición de partituras musicales que se utilizarán para el entrenamiento. Para evaluar nuestro modelo, planeamos calcular resultados cuantitativos en un conjunto de pruebas sintéticas y proporcionar resultados cualitativos en algunas partituras históricas anotadas manualmente.

* + **Detección de símbolos musicales con R-CNN más rápido utilizando anotaciones sintéticas**

La detección precisa de símbolos musicales en imágenes de partituras impresas orquestales o de piano históricas, complejas y densas puede ser un desafío debido a las antiguas técnicas de impresión o la degradación del tiempo. Debido a que los problemas de segmentación pueden variar ampliamente, se necesita un enfoque basado en datos como el uso de detectores de aprendizaje profundo. Sin embargo, la producción de anotaciones de detección (cuadros delimitadores de símbolos + clases) para tales sistemas es costosa y lleva mucho tiempo. Proponemos entrenar dicho modelo con datos sintéticos y anotaciones producidas por un programa de composición tipográfica de música. Analizamos qué clases son relevantes para la tarea de detección y presentamos una primera selección de archivos de composición de partituras musicales que se utilizarán para el entrenamiento. Para evaluar nuestro modelo, planeamos calcular resultados cuantitativos en un conjunto de pruebas sintéticas y proporcionar resultados cualitativos en algunas partituras históricas anotadas anualmente.