

Trabajo de la asignatura

“Aplicación de técnicas de aprendizaje profundo al reconocimiento óptico de partituras SATB”

Aprendizaje Profundo (AP) – Máster Universitario en Inteligencia
Computacional e Internet de las Cosas

Universidad de Córdoba
2022/2023

Trabajo realizado por:
-Antonio Gómez Giménez
(i72gogia@uco.es)

Aplicación de técnicas de aprendizaje profundo al reconocimiento óptico de partituras SATB

Índice :

- Introducción:
- Elección del artículo:
- Problema planteado en el artículo:
- Resolución del problema y metodología implementada:
- Conclusión y posibles mejoras:

Aplicación de técnicas de aprendizaje profundo al reconocimiento óptico de partituras SATB

Índice :

- **Introducción:**
- Elección del artículo:
- Problema planteado en el artículo:
- Resolución del problema y metodología implementada:
- Conclusión y posibles mejoras:

Introducción

- Reforzar los conocimientos obtenidos en la asignatura.
- Elección del artículo que se pretende analizar.
- Se comentarán ciertos aspectos del propio artículo (problema se pretende resolver).
- Resultados.
- Posibles mejoras.

Aplicación de técnicas de aprendizaje profundo al reconocimiento óptico de partituras SATB

Índice :

- Introducción:
- **Elección del artículo:**
- Problema planteado en el artículo:
- Resolución del problema y metodología implementada:
- Conclusión y posibles mejoras:

Elección del artículo:

- Artículo -> **“Aplicación de técnicas de aprendizaje profundo al reconocimiento óptico de partituras SATB”**
- Realizado por :
 - Martin Morita Hernandez y Francisco Fernandez de Vega (Departamento de Tecnología de los Computadores y de las Comunicaciones Universidad de Extremadura Mérida, España)
 - Juan Villegas Cortez (Departamento de Sistemas Universidad Autónoma Metropolitana, Cd. de México, México)
- Razones para la elección:
 - **Origen de la investigación.**
 - **Beneficios para la sociedad.**
 - **Proyección de la misma.**
 - **Gustos personales.**

Aplicación de técnicas de aprendizaje profundo al reconocimiento óptico de partituras SATB

Índice :

- Introducción:
- Elección del artículo:
- **Problema planteado en el artículo:**
- Resolución del problema y metodología implementada:
- Conclusión y posibles mejoras:

Problema planteado en el artículo:

- **Título** -> “Aplicación de técnicas de aprendizaje profundo al reconocimiento óptico de partituras SATB”
- **SABT** -> Soprano, Contralto, Tenor y Bajo.
- **Problema global** -> OMR (Optical Music Recognition)
 - Problemas concretos dentro de este:
 - Armaduras (claves, tempos, etc)
 - Tipo de nota musical
 - Figuras complejas
 - A ordenador/A mano
 - Ejemplos

Problema planteado en el artículo:



Figura 1: Fragmento de Für Elise (Arreglo por Ilsy Sánchez/ Piano Allegro Academia Virtual)

Problema planteado en el artículo:



Figura 2: Interstellar-Suite (Hans Zimmer)

Problema planteado en el artículo:

- **Objetivo** -> Ayudar a *La Sociedad Internacional para la Recuperacion de Informacion Musical* (ISMIR, International Society for Music Information Retrieval)
- **Herramienta de apoyo a Sharpmony**
- **Ayudar a la creación de armonías** -> armonías basadas en armonía clásica (SATB, creadas por Batch)

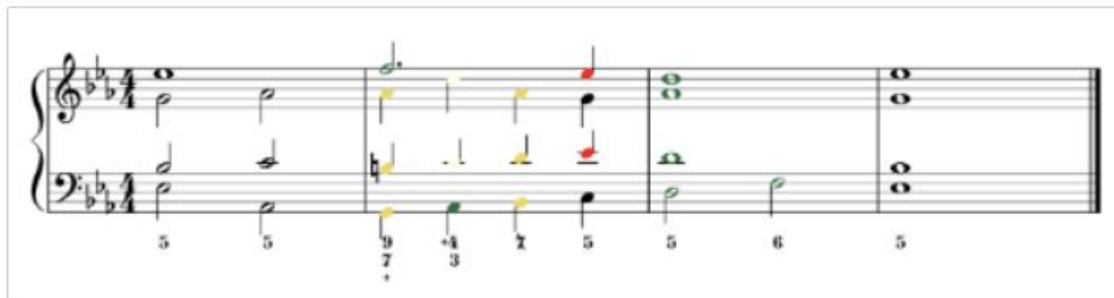


Figura 3: Ejemplo de ejercicio hecho por alumnos de conservatorio (corregido por sharpmony)

Problema planteado en el artículo:

- **Objetivos a lograr para apoyar a ISMIR y como herramienta de Sharpmony:**
 - Corrección automática de ejercicios SATB.
 - Composición automática de corales SATB.
 - Reconocimiento automático de ejercicios SATB escritos a mano (OMR).
- Ejemplo de problema a abordar (del objetivo elegido en este artículo):

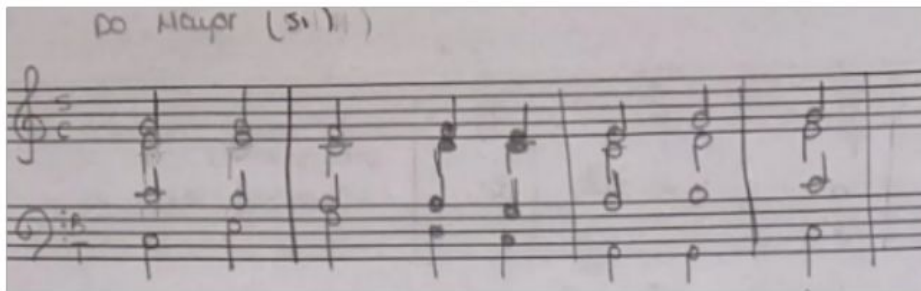


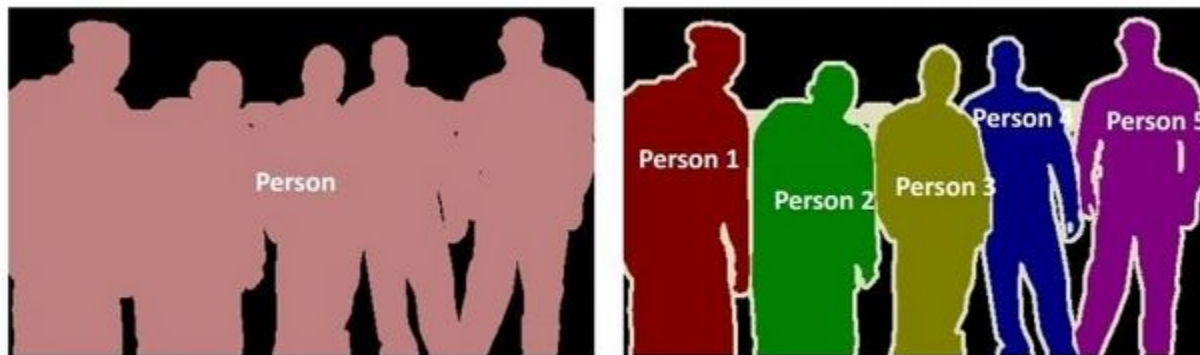
Figura 4: Ejemplo de ejercicio hecho por alumnos de conservatorio a mano en libreta pautada

Problema planteado en el artículo:

- **Enfoque para solucionar el problema:**

- Redes neuronales convolucionales (CNN, Convolutional Neuronal Networks)
 - Mayor velocidad/Menor precisión:
 - YOLO [1]
 - SSD [2]
 - RetinaNet [3]
 - Mayor precisión/Menor velocidad:
 - Faster R-CNN [4]
 - R-CNN [5]
 - R-FCN [6]
- Para este proyecto **Mask R-CNN** [7]

Problema planteado en el artículo:



Semantic Segmentation

Instance Segmentation

Figura 5: Tipos de segmentación aplicada a Mask R-CNN [7]

Aplicación de técnicas de aprendizaje profundo al reconocimiento óptico de partituras SATB

Índice :

- Introducción:
- Elección del artículo:
- Problema planteado en el artículo:
- **Resolución del problema y metodología implementada:**
- Conclusión y posibles mejoras:

Resolución del problema y metodología implementada:

- Estructura general de la metodología que se pretende implementar:

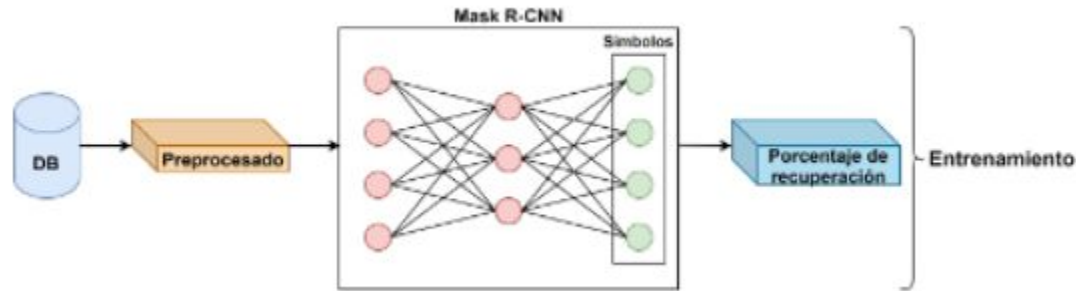


Figura 6: Metodología a implementar

Resolución del problema y metodología implementada:

- **Conjunto de datos (DB).**
 - Fotografías realizadas sobre los cuadernos pautados de los estudiantes de conservatorio (conteniendo partituras SATB)
 - **100 imágenes**
 - **3000 anotaciones a nivel de símbolo**
 - Técnicas de data augmentation para obtener mayor cantidad de patrones
 - Escalado
 - Recorte de imágenes

Resolución del problema y metodología implementada:

- **Preprocesamiento.**

- Primera etapa:
 - Redimensión de imágenes a 1280x720 píxeles.
- Segunda etapa:
 - Especificar figuras a reconocer y el lugar donde se encuentran.
 - Uso de VGG Image Annotator(VIA) [8].
 - Ejemplo de entrada procesada.

Resolución del problema y metodología implementada:

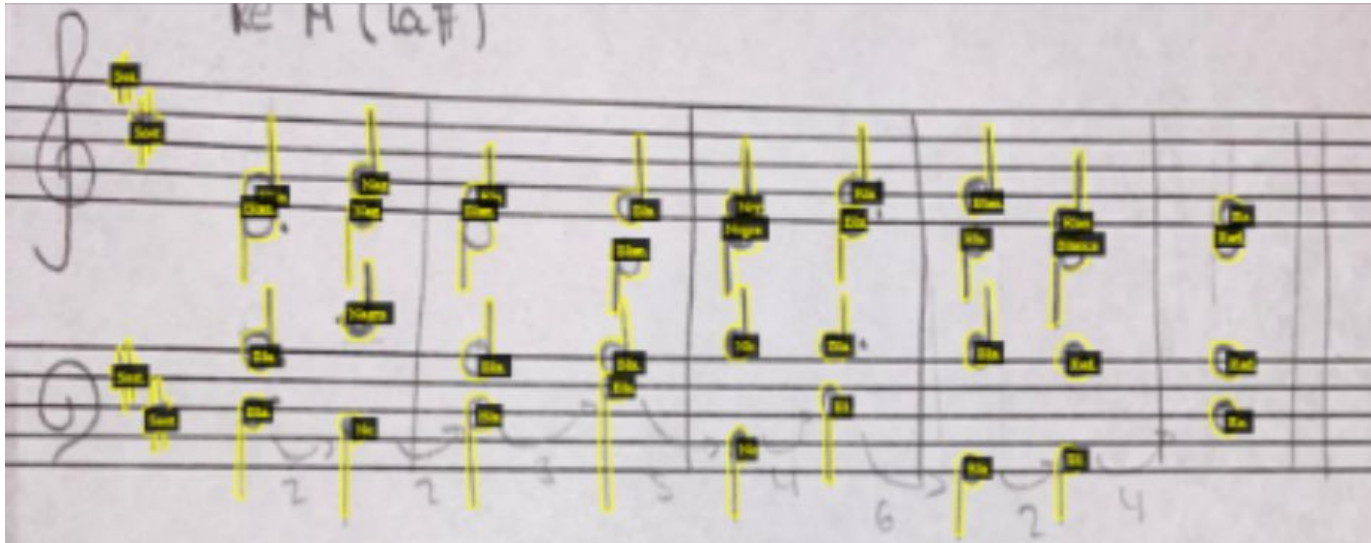


Figura 7: Ejemplo del preprocesamiento del conjunto de datos, para cada partitura SATB

Resolución del problema y metodología implementada:

- **Aplicación de la CNN.**
- **Conceptos de interés:**
 - **ROI (Regiones de interés).** Regiones de interés de la imagen para su clasificación.
 - **RPN (Región propuesta de red).** Red neuronal simple que propone objetos dentro de una imagen
 - **RolPool (Agrupación de regiones de interés).** Para cada cuadro realiza la regresión y clasificación del cuadro delimitador.
 - **Rol Align.** Operación para extraer el pequeño mapa de características de cada ROI de segmentación y clasificación.

Resolución del problema y metodología implementada:

- **Aplicación de la CNN.**
 - Arquitectura Mask R-CNN (presentada en el 2017)

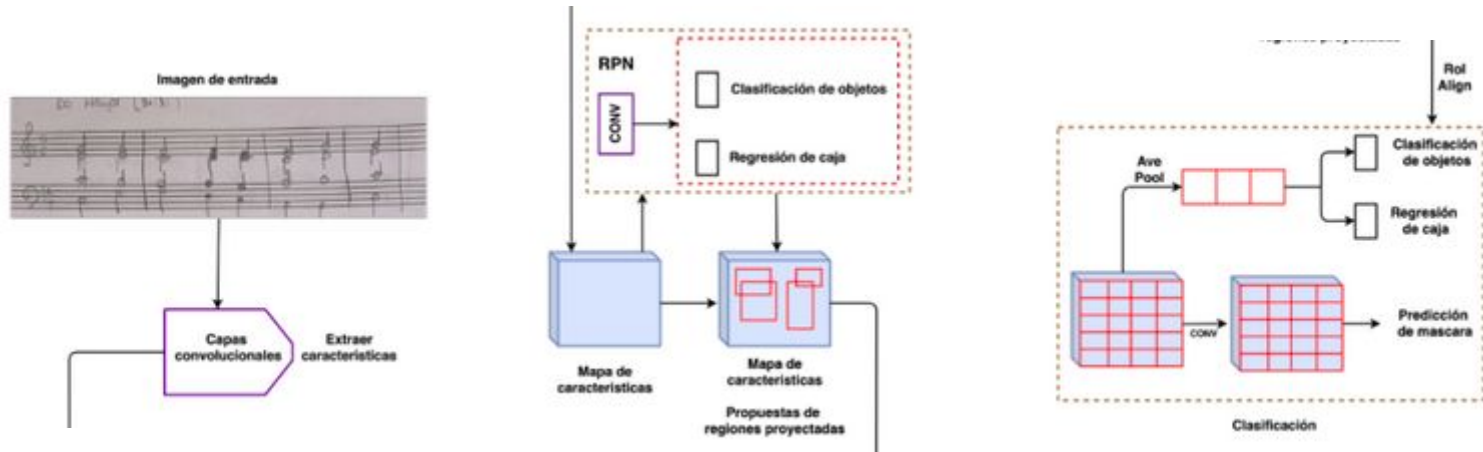


Figura 8: Arquitectura de la red Mask R-CNN

Resolución del problema y metodología implementada:

- **Aplicación de la CNN.**
 - Etapas:
 - **Primera etapa.** Mapa de características para la imagen de entrada.
 - **Segunda etapa.** Las salidas de la primera etapa se utilizan para la red de propuesta de región (RPN) para generar regiones de interés (ROI).
 - **Tercera etapa.** Aprovechando las regiones de interés generadas por el RPN, se mapean y se extraen las características necesarias para realizar las clasificaciones de los objetos, las máscaras de segmentación y los cuadros delimitadores.

Resolución del problema y metodología implementada:

- **Aplicación de la CNN.**
 - Objetos musicales que la red puede predecir:

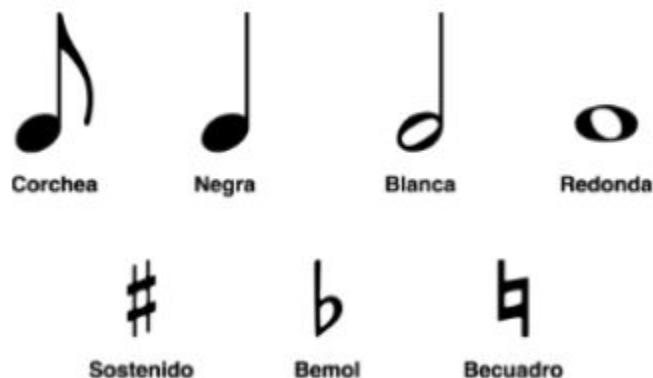


Figura 9: Notas clasificables en las partituras SATB

Resolución del problema y metodología implementada:

- **Porcentaje de recuperación.** (Ajuste de la red)

- La pérdida de clasificación (Lcls).

$$L_{cls}(p_i^*, p_i) = -\log(p_i^* p_i)$$

- Localización (Lbox).

$$L_1^{smooth}(x) = \begin{cases} 0,5x^2 & \text{if } |x| < 1 \\ |x| - 0,5 & \text{en otro caso} \end{cases}$$

$$L_{box}(t_i, t_i^*) = L_1^{smooth}(t_i^* - t_i)$$

Resolución del problema y metodología implementada:

- **Porcentaje de recuperación.** (Ajuste de la red)
 - Máscara de segmentación (Lmask).

$$L_{mask}(s_i, s_i^*) = -(s_i^* \log s_i + (1 - s_i^*) \log(1 - s_i))$$

- Pérdida total del entrenamiento.

$$L(p_i, p_i^*, t_i, t_i^*, s_i, s_i^*) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) \\ + \frac{\lambda}{N_{box}} \sum_i p_i^* L_{box}(t_i, t_i^*) + \frac{\gamma}{N_{mask}} \sum_i L_{mask}(s_i, s_i^*)$$

Resolución del problema y metodología implementada:

- **Porcentaje de recuperación.** (Ajuste de la red)
 - Variables:
 - **p_i** -> representa la probabilidad predicha de que el cuadro delimitador i sea un objeto.
 - **p^*_i** -> representa la probabilidad de verdad básica (binaria) de si el cuadro delimitador i es un objeto.
 - **t_i** -> representa cuatro coordenadas parametrizadas, que son: el valor de las coordenadas horizontales y verticales del punto central en el cuadro, la anchura y la altura del cuadro.
 - **t^*_i** -> indica la diferencia entre el cuadro de la etiqueta verdadera y el cuadro delimitador positivo.
 - **s** -> representan respectivamente las matrices binarias de la máscara de predicción y de la etiqueta verdadera.

Resolución del problema y metodología implementada:

- ¿Cómo medir los resultados?
- Métricas usadas:
 - **precisión (P)**
 - **recuperación (R)**

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

Resolución del problema y metodología implementada:

- Se puede comprobar de forma visual los resultados.

Ejemplo:



Figura 10: Ejemplo de clasificación con imagen nota detectada (derecha) y nota no detectada (izquierda) en partitura SATB

Resolución del problema y metodología implementada:

- Máquina utilizada:
 - NVIDIA GTX 1080 Ti
 - NVIDIA GTX 1080
 - RAM de 64GB
 - Disco duro de 8TB
- Lenguaje utilizado:
 - Python apoyándose en **TenzoFlow** y **Keras**.

Resolución del problema y metodología implementada:

- Resultados obtenidos para:
 - 100 imágenes de partituras SABT
 - 80% entrenamiento
 - 20% validación
 - ¡**Mismas imágenes** para entrenar que para verificar la estabilidad y confiabilidad (test) del modelo entrenado!
 - Parámetros de la red:
 - Tasa de aprendizaje de 0.001
 - Factor de ajuste 0.9%
 - 700 épocas con 100 pasos para cada una

Resolución del problema y metodología implementada:

- Las 100 imágenes de prueba dieron como resultado:
 - **Precisión** ->95.4%
 - **Recuperación** -> 94.5%
- Ejemplo de salida visual:

Resolución del problema y metodología implementada:



Figura 10: Ejemplo de salida (predicción)

Aplicación de técnicas de aprendizaje profundo al reconocimiento óptico de partituras SATB

Índice :

- Introducción:
- Elección del artículo:
- Problema planteado en el artículo:
- Resolución del problema y metodología implementada:
- **Conclusión y posibles mejoras:**

Resolución del problema y metodología implementada:

- **¿Se cumplieron los objetivos del artículo?**
 - Parcialmente
- **Problemas:**
 - Imbalance en las clases
 - Redondas -> 25%
 - Blancas -> 40%
 - Negras -> 20%
 - Corcheas, sostenidos, bemoles y becuadros -> 2%, 6%, 5% y 2%
 - Problemas de etiquetado

Resolución del problema y metodología implementada:

- **Posibles soluciones:**
 - Arreglar el problema de la clase minoritaria
 - Muestras artificiales
 - Modificar el Dataset
 - Ajustar los parámetros del modelo

Resolución del problema y metodología implementada:

- **Conclusiones generales:**
 - Tema interesante y con proyección
 - Objetivo del proyecto
 - Objetivo general de reconocimiento de partituras
- **Conclusiones del trabajo:**
 - Aprendizaje de nuevas redes
 - Aplicaciones de cnn a problemas reales

Bibliografía:

- [1] J. Redmon, S. Divvala, R. Girshick, y A. Farhadi, You only look once: Unified, real-time object detection, en Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779-788.
- [2] W. Liu et al., Ssd: Single shot multibox detector, en European conference on computer vision, 2016, pp. 21-37.
- [3] T.-Y. Lin, P. Goyal, R. Girshick, K. He, y P. Dollar, ' Focal loss for dense object detection, en Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980-2988.
- [4] S. Ren, K. He, R. Girshick, y J. Sun, Faster r-cnn: Towards realtime object detection with region proposal networks, arXiv preprint arXiv:1506.01497, 2015.
- [5] R. Girshick, J. Donahue, T. Darrell, y J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, en Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580-587.
- [6] J. Dai, Y. Li, K. He, y J. Sun, R-fcn: Object detection via region-based fully convolutional networks, arXiv preprint arXiv:1605.06409, 2016.
- [7] Odemakinde, E. (2022, 11 julio). Everything about Mask R-CNN: A Beginner's Guide. viso.ai. <https://viso.ai/deep-learning/mask-r-cnn>

Bibliografía:

[8] Zhang, Y., Chu, J., Leng, L., & Miao, J. (2020). Mask-Refined R-CNN: A Network for Refining Object Details in Instance Segmentation. *Sensors*, 20(4), 1010. <https://doi.org/10.3390/s20041010>



¡Muchas gracias por su
atención!

Ejemplo más detallado de la arquitectura Mask R-CNN

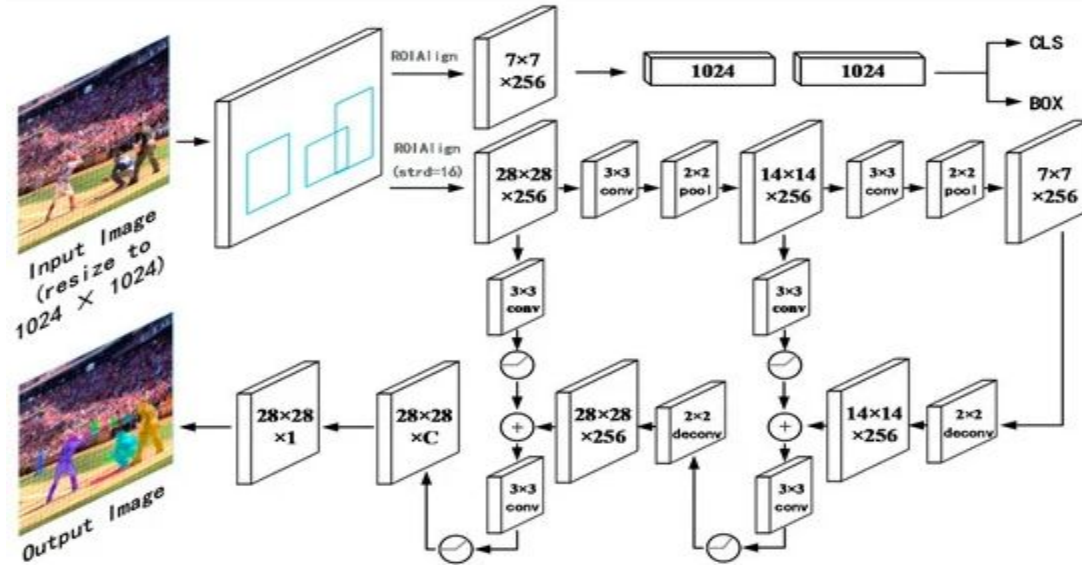


Figura 11: Arquitectura Mask R-CNN.