



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Dpto. de Comunicaciones

Estudio bibliográfico sobre la compresión de vídeo y el uso de IA y la evaluación de prestaciones de una solución propuesta.

Trabajo Fin de Máster

Máster Universitario en Tecnologías, Sistemas y Redes de Comunicaciones

AUTOR/A: Chasi Sandoval, Erik Joel

Tutor/a: Guerri Cebollada, Juan Carlos

Cotutor/a: Arce Vila, Pau

CURSO ACADÉMICO: 2023/2024

Objetivos —

- Analizar el estado del arte respecto a la codificación de vídeo y el uso de IA, mediante un estudio bibliográfico con base en fuentes y repositorios científicos para extraer datos de carácter relevante para el campo de la codificación de vídeo.
- Evaluar las prestaciones de codificadores de vídeo IA a través de métricas de calidad objetiva y perceptual para realizar una comparativa con las prestaciones ofrecidas por la codificación tradicional de los códecs H.264 y H.265.
- Discutir los resultados más relevantes encontrados en el estudio bibliográfico y la evaluación de prestaciones de codificadores de vídeo IA para determinar los desafíos y líneas de investigación para el campo de la codificación de vídeo.

Metodología — La metodología planteada para abordar el desarrollo del presente trabajo se ejecuta a través de cuatro etapas o fases que comienza con una introducción al tema donde se abordan las generalidades, importancia y aplicaciones de la codificación de vídeo; la segunda etapa corresponde al estudio bibliográfico que empieza con la búsqueda bibliográfica sobre la cual se realiza una interpretación a través de un análisis cualitativo y cuantitativo; la tercera etapa está destinada a la evaluación de prestaciones de codificadores de vídeo IA que se lleva a cabo a través de las fases de pre-codificación, codificación y post-codificación. Finalmente se realiza una etapa de discusión donde se exponen los datos más relevantes de la etapa del estudio bibliográfico y la evaluación de prestaciones abordando los principales desafíos y líneas de investigación en el campo de la codificación de vídeo.

Desarrollos teóricos realizados — El desarrollo teórico de este trabajo se aborda en las dos primeras etapas que describe la metodología donde se habla acerca de las generalidades, importancia y aplicaciones de la codificación de vídeo, posterior a lo cual se presentan las fuentes y repositorios de búsqueda empleadas, así como las herramientas a utilizar para analizar e interpretar los datos de los artículos científicos que se tomaran como sustento para este estudio, en función de los datos recopilados se desarrolla un análisis cuantitativo a través de diagramas aluviales, mapas de proyección y redes bibliométricas para luego llevar a cabo un análisis cualitativo donde se habla acerca de la evolución de la codificación de vídeo, el enfoque tradicional y el enfoque IA para la codificación de vídeo, las métricas para evaluar prestaciones y finalmente las publicaciones que aportan con código abierto para estudiar y evaluar los modelos de codificadores basados en IA.

Desarrollo de prototipos y trabajo de laboratorio — El trabajo de laboratorio se desarrolla en la tercera etapa de la metodología, su finalidad es evaluar las prestaciones de los codificadores de vídeo con IA, comienza con un análisis pre-codificación donde se seleccionan los archivos multimedia que conforman el banco de vídeos de prueba sobre los cuales se extraen los parámetros característicos como la información general y la información de la complejidad espacial y temporal (SITI), luego se realiza la fase de codificación sobre los vídeos de prueba utilizando los códecs tradicionales (H.264/H.265) y los códecs basados en IA (ENVC/ALVC). Finalmente se ejecuta un análisis post-codificación que emplea las métricas de calidad de vídeo objetivas (PSNR) y calidad de vídeo perceptual (VMAF) para obtener gráficas con los valores encontrados para su posterior interpretación y discusión.



Resultados — Los resultados del estudio bibliográfico muestran una constante evolución del campo de la codificación de vídeo con IA, desde la aparición del primer estándar de códec de vídeo (H.120) en 1984 hasta la utilización del aprendizaje profundo y las redes neuronales para el desarrollo de herramientas y códecs en la actualidad. La evaluación de las prestaciones de las soluciones propuestas muestra que la codificación tradicional presenta mayor eficiencia en términos de calidad visual con mejores medidas de PSNR, sin embargo, la codificación IA presenta mayor eficiencia en términos de compresión al reducir significativamente el bitrate real del vídeo codificado. Las medidas de VMAF muestran que ambas soluciones aportan con una buena calidad de vídeo, en bitrates bajos existen ligeras degradaciones perceptibles pero aceptables para los usuarios que no afectan a la experiencia visual y en los bitrates altos se observa una excelente calidad de vídeo.

Líneas futuras — La investigación muestra que las líneas futuras de investigación para la codificación de vídeo con IA están enfocadas en la reducción de la complejidad computacional para una compresión eficiente, el desarrollo de esquemas completos de codificación sobre redes neuronales profundas (DNN) y algoritmos de codificación IA para superresolución de vídeo.

Publicaciones — Hasta el momento no se han realizado publicaciones.

Abstract — The objective of this master's thesis was to conduct a bibliographic study on video coding and the use of AI, along with the performance evaluation of proposed solutions. First, a bibliographic study was carried out with a qualitative and quantitative analysis of publications sourced from the ACM, ARXIV, IEEE Xplore, Science Direct, and Springer repositories from 2019 to 2024. The analysis involved the use of alluvial diagrams, projection maps, tree maps, bar charts, and bibliometric networks to examine parameters such as author, year, source, number of views, number of citations, keywords, title and abstract terms, geographical region, and publications with open-source code on GitHub. Secondly, the performance evaluation was conducted in three stages. The first stage involved a pre-coding analysis where a test video dataset consisting of four videos was selected, from which general information and spatial (SI) and temporal (TI) complexity information were extracted. The second stage involved video coding using H.264/H.265 (traditional coding) and ENVC/ALVC (AI coding). The third stage focused on post-coding, where measurements for objective video quality evaluation metrics (PSNR) and perceptual video quality evaluation metrics (VMAF) were obtained. Finally, a discussion was held on the relevant data from the bibliographic study, a review of the performance evaluation, and future research directions.

Autor: Chasi Sandoval Erik Joel, email: echasi@teleco.upv.es

Director 1: Guerri Cebollada Juan Carlos, email: jcguerri@dcom.upv.es

Director 2: Arce Vila Pau, email: paarvi@iteam.upv.es

Fecha de entrega: 02-09-2024

ÍNDICE

I.	INTRODUCCIÓN.....	5
II.	ESTUDIO BIBLIOGRÁFICO	6
II.1.	BÚSQUEDA BIBLIOGRÁFICA.....	6
II.1.1.	Fuentes y repositorios de búsqueda.....	6
II.1.2.	Herramientas para análisis de datos	8
II.2.	ANÁLISIS CUANTITATIVO	8
II.2.1.	Análisis por autor, año, fuente, visualizaciones y citas.....	9
II.2.2.	Análisis por año de publicación	12
II.2.3.	Análisis por región geográfica	12
II.2.4.	Análisis bibliométrico por autor.....	14
II.2.5.	Análisis bibliométrico por palabras clave	16
II.2.6.	Análisis bibliométrico por título y resumen.....	18
II.3.	ANÁLISIS CUALITATIVO	20
II.3.1.	Evolución de la codificación de vídeo	20
II.3.2.	Enfoque tradicional para la codificación de vídeo.....	21
II.3.3.	Enfoque IA para la codificación de vídeo.....	22
II.3.4.	Métricas para evaluación de prestaciones en la codificación de vídeo	23
II.3.5.	Publicaciones con código abierto en GitHub para la codificación de vídeo	24
III.	EVALUACIÓN DE PRESTACIONES	25
III.1.	ANÁLISIS PRE-CODIFICACIÓN.....	25
III.1.1.	Información general banco de vídeos de prueba.....	25
III.1.2.	Información de la complejidad espacial y temporal (SITI).....	25
III.2.	ANÁLISIS POST-CODIFICACIÓN	25
III.2.1.	Valores PSNR GoP 12 codificación frames tipo IP	26
III.2.2.	Valores PSNR GoP 24 codificación frames tipo IP	27
III.2.3.	Valores PSNR GoP 12 codificación frames tipo IPB	28
III.2.4.	Valores PSNR GoP 24 codificación frames tipo IPB	29
III.2.5.	Valores VMAF GoP 12 codificación frames tipo IPB.....	30
III.2.6.	Valores VMAF GoP 24 codificación frames tipo IPB.....	31
IV.	DISCUSIÓN	32
IV.1.	DATOS RELEVANTES DEL ESTUDIO BIBLIOGRÁFICO	32
IV.2.	REVISIÓN DE LA EVALUACIÓN DE PRESTACIONES	34
IV.3.	LINEAS FUTURAS DE INVESTIGACIÓN	36
V.	CONCLUSIONES.....	37
	AGRADECIMIENTOS	38
	BIBLIOGRAFÍA BASE	38



ÍNDICE DE TABLAS

Tabla 1: Repositorios bibliográficos y artículos científicos.....	8
Tabla 2: Publicaciones código abierto GitHub.....	24
Tabla 3: Información general vídeos de prueba.....	25
Tabla 4: Información complejidad espacial y temporal (SITI) vídeos de prueba.....	25
Tabla 5: Valores PSNR GoP 12 codificación frames tipo IP.....	26
Tabla 6: Valores PSNR GoP 24 codificación frames tipo IP.....	27
Tabla 7: Valores PSNR GoP 12 codificación frames tipo IPB.....	28
Tabla 8: Valores PSNR GoP 24 codificación frames tipo IPB.....	29
Tabla 9: Valores VMAF GoP 12 codificación frames tipo IPB.....	30
Tabla 10: Valores VMAF GoP 24 codificación frames tipo IPB.....	31

ÍNDICE DE FIGURAS

Fig. 1: Esquema metodología.....	6
Fig. 2: Mapa de árbol artículos totales por repositorio.....	7
Fig. 3: Diagrama aluvial estudio de publicaciones repositorio ACM.....	9
Fig. 4: Diagrama aluvial estudio de publicaciones repositorio ARXIV.....	9
Fig. 5: Diagrama aluvial estudio de publicaciones repositorio IEEE.....	10
Fig. 6: Diagrama aluvial estudio de publicaciones repositorio Science Direct.....	11
Fig. 7: Diagrama aluvial estudio de publicaciones repositorio Springer.....	11
Fig. 8: Diagrama de barras total de publicaciones por año y repositorio.....	12
Fig. 9: Mapa de proyección ACM/ARXIV/IEEE/Science Direct/Springer.....	13
Fig. 10: Mapa de proyección publicaciones total de repositorios.....	14
Fig. 11: Red bibliométrica por autor ACM/ARXIV/IEEE/Science Direct/Springer.....	15
Fig. 12: Red bibliométrica por autor publicaciones total de repositorios.....	16
Fig. 13: Red bibliométrica por palabras clave ACM/ARXIV/IEEE/Science Direct/Springer.....	17
Fig. 14: Red bibliométrica por palabras clave publicaciones total de repositorios.....	18
Fig. 15: Red bibliométrica por título y resumen ACM/ARXIV/IEEE/Science Direct/Springer.....	19
Fig. 16: Red bibliométrica por título y resumen publicaciones total de repositorios.....	20
Fig. 17: Diagrama línea de tiempo evolución codificación de vídeo.....	21
Fig. 18: Enfoque codificación vídeo con IA.....	22
Fig. 19: Gráfica complejidad espacial y temporal (SITI) vídeos de prueba.....	25
Fig. 20: Gráficas PSNR GoP 12 frames IP.....	26
Fig. 21: Gráficas PSNR GoP 24 frames IP.....	27
Fig. 22: Gráficas PSNR GoP 12 frames IPB.....	28
Fig. 23: Gráficas PSNR GoP 24 frames IPB.....	29
Fig. 24: Gráficas VMAF GoP 12 frames IPB.....	30
Fig. 25: Gráficas VMAF GoP 24 frames IPB.....	31

I. INTRODUCCIÓN

La codificación de vídeo busca principalmente optimizar el archivo de vídeo reduciéndolo a través de la compresión sin perder la calidad visual para asegurar la experiencia para el usuario, esto es fundamental para aplicaciones que requieren transmisión de contenido multimedia, almacenamiento de contenido en dispositivos o medios con espacio limitado y en el área de comunicaciones para transmisiones de vídeo. En los últimos años la compresión de vídeo ha tenido relevancia y ha experimentado una constante evolución, esto se debe en gran medida a la aparición nuevas formas de entretenimiento como las plataformas que ofrece streaming de vídeo, aplicaciones de videollamada que brindan opciones de conexión y comunicación a los usuarios a nivel global, contenido multimedia en vivo como transmisiones de televisión o en directo, almacenamiento en la nube y redes sociales donde se consumen gran cantidad de contenido multimedia.

Los sistemas de compresión de propósito general pueden ser de dos tipos: con pérdidas o sin pérdidas. La compresión sin pérdidas persigue comprimir el archivo de vídeo sin perder calidad del mismo, es decir que los datos originales se pueden recuperar a partir de los datos comprimidos. Por otra parte, la compresión con pérdidas busca reducir el tamaño del archivo a cambio de reducir la calidad del vídeo. Algunos de los parámetros configurables en la codificación de vídeo son la resolución, la cantidad de bits por segundo o bitrate, el número de cuadros o frames por segundo (fps), la estructura del grupo de imágenes o GoP, entre otros.

Los archivos multimedia de vídeo están compuestos por una secuencia de imágenes o frames. En la codificación de vídeo se distinguen tres tipos: los frames I (*Intra*) son independientes de otros frames para su codificación, los frames P (*Predicted*) requieren de frames anteriores para su codificación y finalmente los frames B (*Bidirectional*) que necesitan de frames anteriores y siguientes para su codificación. El conjunto de un número determinado de estos frames forman una estructura de grupo de imágenes o GoP que dentro de un códec de vídeo es determinante para mantener un equilibrio entre la calidad de vídeo y el nivel de compresión.

Los códec de vídeo buscan una compresión de vídeo eficiente sin comprometer la calidad visual y para esto utilizan modelos de codificación temporal, espacial y entrópica. El modelo de codificación temporal utiliza los vectores de movimiento para reducir redundancias a través de la predicción y compensación de movimiento; por otro lado, el modelo de codificación espacial se vale de transformaciones y la cuantización de los coeficientes de transformación para reducir redundancias; por último, el modelo de codificación entrópica se aplica luego de la codificación temporal y espacial y utiliza algoritmos que buscan la representación de símbolos frecuentes con menos bits a través de secuencias codificadas que ocupan menos espacio.

El presente trabajo tiene como fin realizar un estudio bibliográfico sobre la compresión de vídeo y el uso de la Inteligencia Artificial (IA), así como la evaluación de las prestaciones que puedan brindarnos las nuevas soluciones propuestas de códec o modelos con IA. Para llevar a cabo a esta

tarea se propone la organización del trabajo a través de etapas o fases que cubren la metodología de este y se muestran de manera detalla en la Figura 1. La finalidad de este trabajo es explorar el estado del arte de la codificación de vídeo con IA y discutir los principales hallazgos de las prestaciones que pueden brindarnos los nuevos modelos que se encuentran desarrollando y estudiando actualmente.

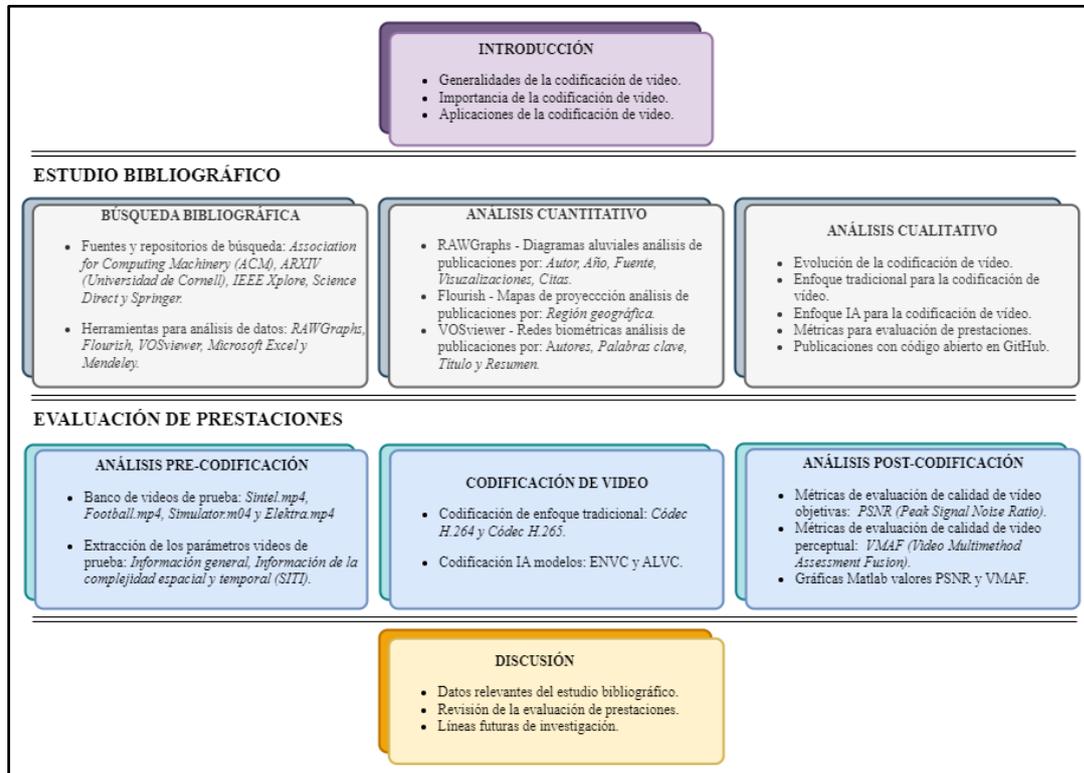


Fig. 1: Esquema metodología.

II. ESTUDIO BIBLIOGRÁFICO

II.1. BÚSQUEDA BIBLIOGRÁFICA

En esta sección se presenta los recursos y herramientas para llevar a cabo la búsqueda bibliográfica de este trabajo que permite identificar, evaluar, extraer y sintetizar información o datos de carácter relevante para este estudio. Primero se describen las fuentes y repositorios que contienen los artículos científicos que fueron considerados para el estudio, así como los criterios de búsqueda. Y luego se presentan las herramientas para analizar e interpretar los datos a través de diagramas aluviales, mapas de protección, mapas de árbol, diagramas de barras y redes bibliométricas que toman parámetros como los autores, palabras clave, títulos y resumen, región geográfica, años de publicación, visualizaciones, citas, entre otros.

II.1.1. Fuentes y repositorios de búsqueda

Para el desarrollo de este trabajo se tomaron como base de búsqueda 5 repositorios de carácter científico donde se publican divulgaciones de distintas disciplinas científicas y técnicas, entre ellas la *multimedia* y *comunicación*. Los repositorios seleccionados fueron: Association for Computing Machinery (ACM), Universidad de Cornell (ARXIV), IEEE Xplore, Science Direct y Springer.

La Figura 2 muestra el total de artículos tomados por repositorio bibliográfico, siendo un total de 115. Como se puede observar, IEEE Xplore presenta la gran mayoría de artículos 58 en total, esto se debe a que contienen publicaciones especializadas en investigación tecnológica de las áreas de la ingeniería eléctrica, electrónica, telecomunicaciones e informática; el segundo lugar es para Springer con 23 artículos, este repositorio se caracteriza por ser una editorial multidisciplinar que abarca una gran cantidad de disciplinas y áreas de estudio; el tercer lugar lo ocupa ACM con 14 artículos y este se caracteriza por promover la ciencia y la educación en informática por medio de publicaciones de revistas y conferencias; el cuarto lugar lo tiene ARXIV con 12 artículos, este repositorio gratuito permite acceder a contenido científico antes de ser publicado oficialmente por revistas académicas y cubre áreas de interés como ciencias de la computación e ingeniería; por último, encontramos a Science Direct con 8 artículos, esta es una plataforma que permite acceder a artículos científicos y técnicos de Elsevier.

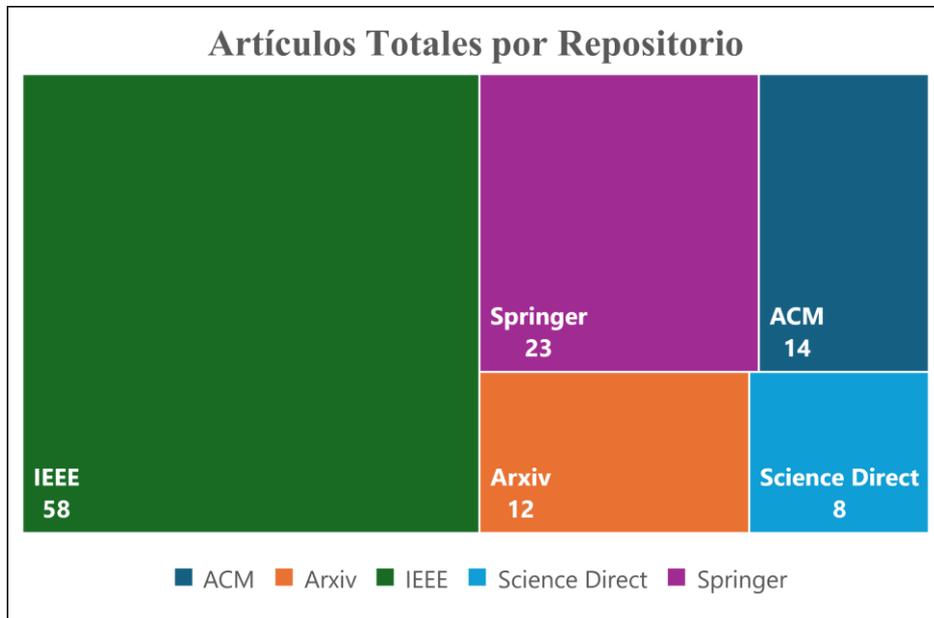


Fig. 2: Mapa de árbol artículos totales por repositorio.

Los motores de búsqueda de estos repositorios toman como base una serie de parámetros de entrada generales y comunes en todos ellos para filtrar resultados con similitud o coincidencia de lo que se requiere buscar. En este caso se consideraron como palabras de búsqueda los términos: *codificación de vídeo*, *inteligencia artificial* y *compresión de vídeo*, como rango de los años de publicación se consideró desde *2019 a 2024*, finalmente las áreas de interés consideradas fueron: *sistemas de información multimedia*, *comunicaciones*, *computación e inteligencia artificial*.

Los artículos que fueron considerados para este estudio bibliográfico se presentan clasificados por repositorio en la Tabla 1 para su posterior análisis, tomando en consideración la información de publicación que se puede extraer de ellos como son autores, año de publicación, fuente, visualizaciones, citas, región geográfica, palabras clave, título y resumen, además de aquellos que son publicaciones con código abierto que permita acceder a modelos de prueba y estudio.

REPOSITORIO	ARTÍCULOS CIENTÍFICOS
ACM	[1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14]
ARXIV	[15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26]
IEEE	[27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73], [74], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84]
SCIENCE DIRECT	[85], [86], [87], [88], [89], [90], [91], [92]
SPRINGER	[93], [94], [95], [96], [97], [98], [99], [100], [101], [102], [103], [104], [105], [106], [107], [108], [109], [110], [111], [112], [113], [114], [115]

Tabla 1: Repositorios bibliográficos y artículos científicos.

II.1.2. Herramientas para análisis de datos

Los artículos científicos manejan grandes volúmenes de datos, pero también poseen parámetros que se pueden interpretar y analizar si se representan gráficamente. En este trabajo los campos de análisis que serán presentados por medio de diagramas, mapas y redes son: autores del artículo, año de publicación, fuente de publicación del artículo, número de visualizaciones, citas realizadas a los artículos, región geográfica de procedencia, palabras clave, título y resumen del artículo, y publicaciones con código abierto disponible.

El análisis e interpretación de los campos: autores, año de publicación, fuente de publicación, visualizaciones y citas realizadas al artículo se interpretan a través de diagramas aluviales por cada uno de los repositorios. La herramienta que nos permite realizar esta representación gráfica es RAWGraphs, esta es una herramienta de visualización de código abierto que combina hojas de cálculo que almacenan los datos a interpretar con los editores gráficos vectoriales, de esta manera se logra una representación visual de datos complejos de manera sencilla [116].

El análisis del total de publicaciones que son consideradas para este trabajo por cada uno de los repositorios y publicadas en el rango de años comprendido entre 2019 a 2024, se puede visualizar de mejor manera a través de diagramas de barras realizadas con el software de hojas de cálculo de Microsoft-Excel. Para interpretar el campo de región geográfica de procedencia de las publicaciones se considera adecuado emplear mapas de proyección realizados en Flourish que permite crear elementos visuales de datos organizados en hojas de cálculo [117].

Los campos característicos de un artículo científico son por lo general: los autores, palabras clave, título y resumen, las conexiones, relaciones, similitudes y concurrencias entre cada uno de ellos y se puede representar de manera gráfica a través de redes bibliométricas que pueden trazarse utilizando herramientas como VOSviewer que nos permite construir redes en función de relaciones de citas, acoplamiento bibliográfico, cocitación o coautoría y minería de texto científica [118], [119].

II.2. ANÁLISIS CUANTITATIVO

La sección dedicada al análisis cuantitativo busca explorar las relaciones que existen entre los artículos científicos que hablan acerca de la codificación de vídeo con IA que son el objeto de nuestra investigación. Se pretende identificar aquellos artículos que destacan por su número de

visualizaciones o citas, así como estudiar el índice de publicaciones de este tema en los últimos años y analizar la tendencia de publicaciones sobre este campo de investigación por región geográfica. Por último, se representan las relaciones de concurrencia por autor, palabras clave, título y resumen a través de estructuras de redes bibliométricas para determinar los más importantes.

II.2.1. Análisis por autor, año, fuente, visualizaciones y citas

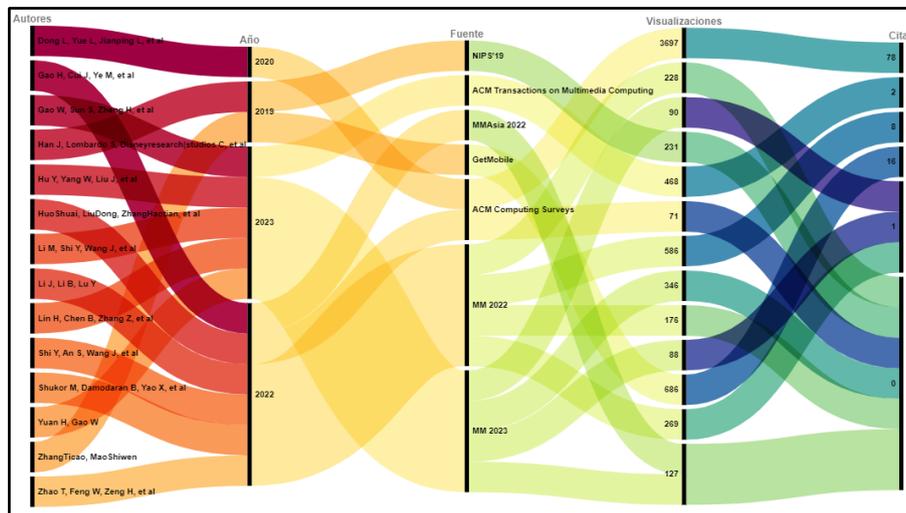


Fig. 3: Diagrama aluvial estudio de publicaciones repositorio ACM.

El repositorio ACM cuenta con 14 publicaciones en total sobre este tema, la Figura 3 muestra el diagrama aluvial que representa el flujo de cada una de estas publicaciones. Se puede observar una fuerte actividad de publicación en los años 2022 y 2023 con un total de 10 cuya fuente de publicación es la 30ª y 31ª edición de la “*Conferencia Internacional ACM sobre Multimedia*”, destacar la publicación de “*Deep Learning-Based Video Coding*” - Dong L, Yue L, Jianping L, et al [14] que cuenta con 3697 visualizaciones y es la publicación más citada con un total de 78, seguida de “*An Overview of Emerging Video Coding Standards*” - Zhang Ticao & Mao Shiwen [13] con 686 visualizaciones y 16 citaciones.

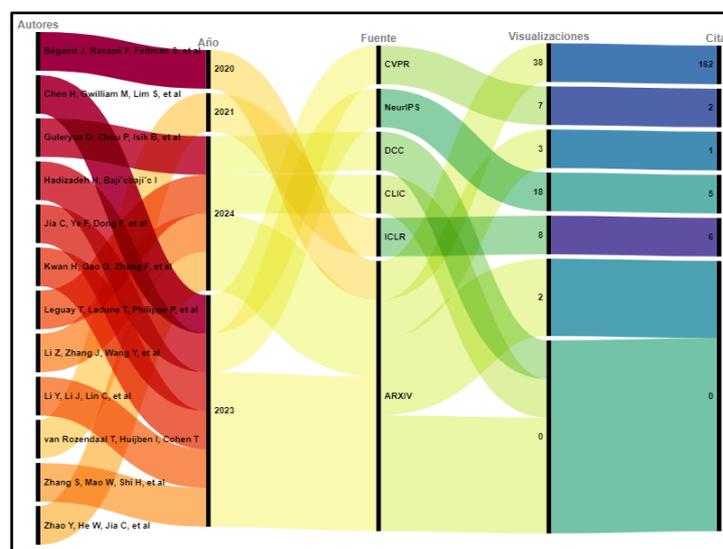


Fig. 4: Diagrama aluvial estudio de publicaciones repositorio ARXIV.

El repositorio ARXIV cuenta con 12 publicaciones, la Figura 4 muestra el diagrama aluvial con el flujo de cada una de ellas. Se puede observar una fuerte actividad de publicación para los años 2023 y 2024 con un total de 10, la publicación relevante de este repositorio es “*CompressAI: a PyTorch library and evaluation platform for end-to-end compression research*” - Bégaint J, Racapé F, Feltman S, et al [26] que cuenta con 38 visualizaciones y a su vez es la publicación más citada con 162 en total; seguida tenemos la publicación “*HiNeRV: Video Compression with Hierarchical Encoding-based Neural Representation*” - Kwan H, Gao G, Zhang F, et al [16] publicada en la “*Conference on Neural Information Processing*” con 18 visualizaciones y 5 citas en total.

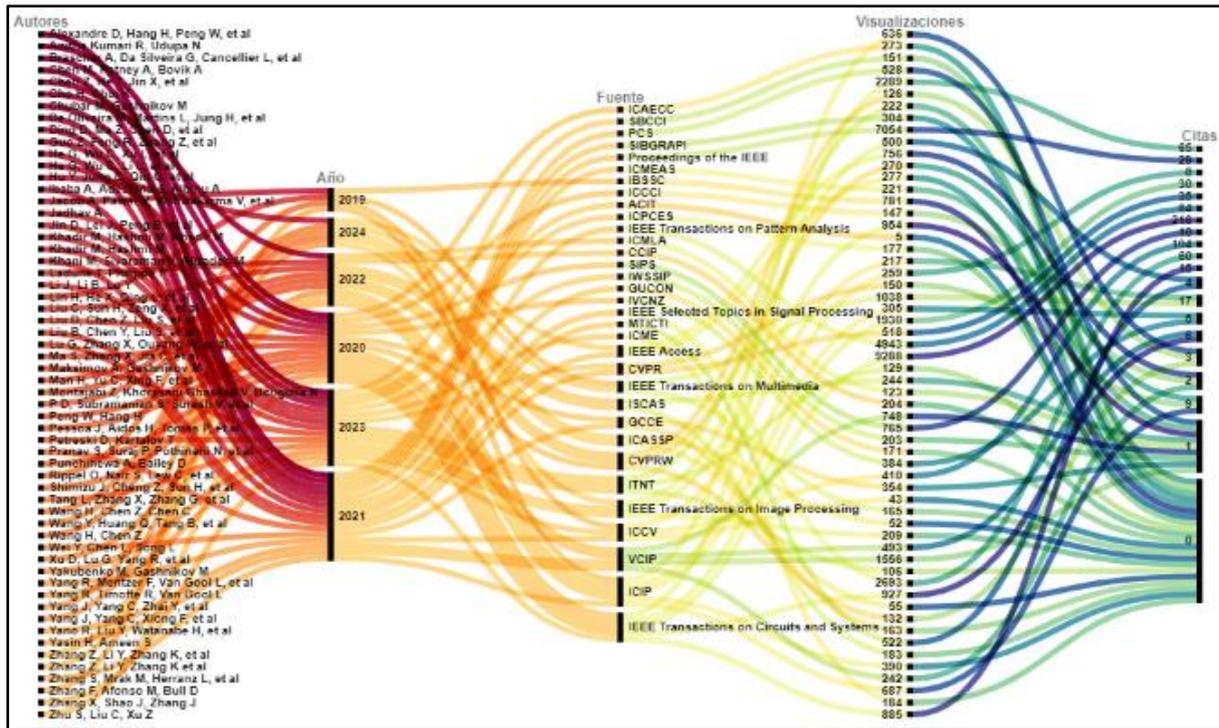


Fig. 5: Diagrama aluvial estudio de publicaciones repositorio IEEE.

El repositorio IEEE aporta 58 publicaciones, la Figura 5 muestra el diagrama aluvial con el flujo de cada una de ellas. Aquí se observa que los años 2020, 2021 y 2023 tiene una fuerte actividad de publicación con un total de 40, la fuente principal de publicación de manera general son simposios y conferencias pero también se destaca la revista científica “*IEEE Transactions on Circuits and Systems*” que cubre temas de procesamiento de señal, se destaca “*Image and Video Compression With Neural Networks: A Review*” – Ma S, Zhang X, Jia C, et al [71] como la publicación con el mayor número de visualizaciones con 9288 en total y su vez es la más citada con 216, seguido tenemos la publicación “*Advances in Video Compression System Using Deep Neural Network: A Review and Case Studies*” - Ding D, Ma Z, Chen D, et al [32] con 7054 visualizaciones y 29 citas, luego encontramos a “*An End-to-End Learning Framework for Video Compression*” - Lu G, Zhang X, Ouyang W, et al [70] con 4943 visualizaciones y 84 citas.

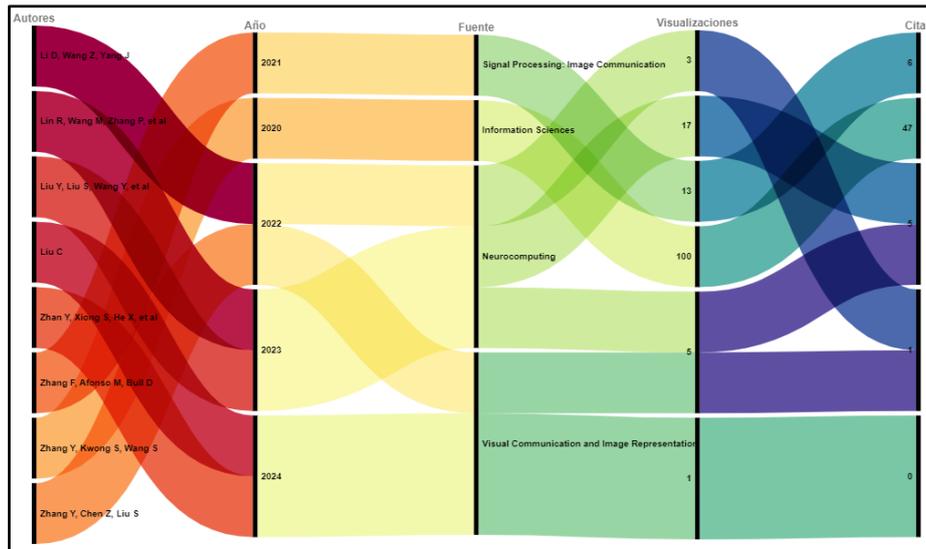


Fig. 6: Diagrama aluvial estudio de publicaciones repositorio Science Direct.

Science Direct cuenta con 8 publicaciones, la Figura 6 muestra el diagrama aluvial con el flujo de cada una de ellas. Se puede observar que la mayor actividad de publicación está en los años 2022 a 2024 con un total de 6, la publicación relevante de este repositorio es “Machine learning based video coding optimizations: A survey” - Zhang Y, Kwong S, Wang S [90] que cuenta con 100 visualizaciones y a su vez es la publicación más citada con 47 en total, las revistas científicas “Neurocomputing” y “Visual Communication and Image Representation” donde se presentan temas de vanguardia en inteligencia artificial, aprendizaje automático, computación neural, imágenes y comunicaciones abarcan el mayor número de publicaciones.

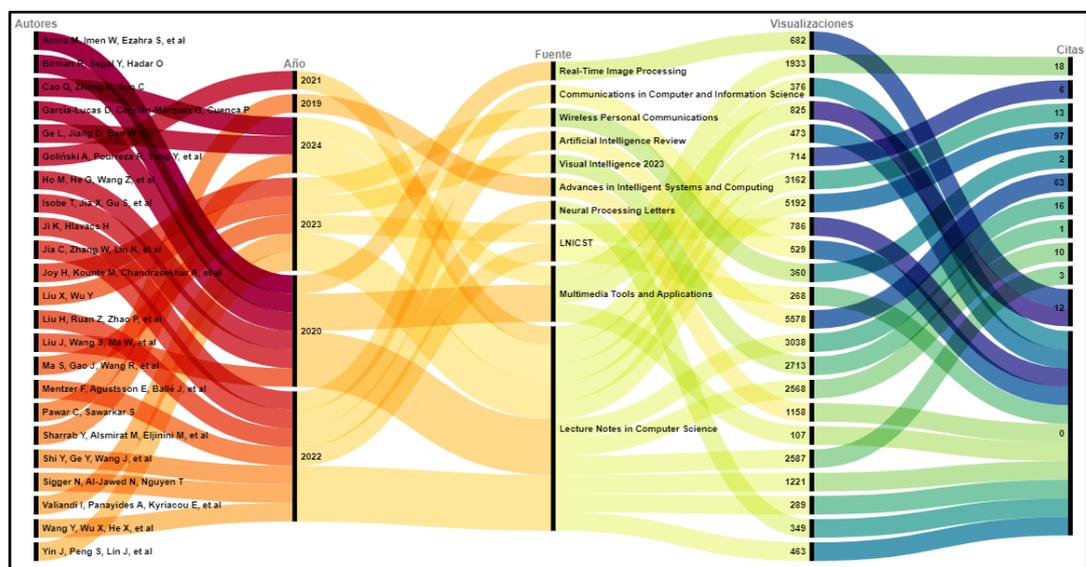


Fig. 7: Diagrama aluvial estudio de publicaciones repositorio Springer.

Springer cuenta con 23 publicaciones, la Figura 7 muestra el diagrama aluvial con el flujo de cada una de ellas y se puede observar que la mayor actividad de publicación está en los años 2020, 2022 y 2023 con un total de 18. La publicación relevante de este repositorio es “Video super-resolution based on deep learning: a comprehensive survey” - Liu H, Ruan Z, Zhao P, et al [106] que cuenta

con 5578 visualizaciones y 63 citas, luego tenemos a “*Vídeo Super-Resolution with Recurrent Structure-Detail Network*” - Isobe T, Jia X, Gu S, et al [107] que tiene 5192 visualizaciones y el mayor número de citaciones de este repositorio con 97, además forma parte de serie de actas de congresos “*Lecture Notes in Computer Science (LNCS)*” que destaca como fuente de publicación de este repositorio.

II.2.2. Análisis por año de publicación

Conocer el volumen de publicaciones que se realizan cada año sobre este tema es fundamental para tener una idea de la tendencia de investigación que se tendrá para los próximos años en este campo. Como se observa en la Figura 8, 2023 es el año con más publicaciones con un total de 31 sumando todos los repositorios, seguido de 2022 con 24 publicaciones y 2020 con 21 publicaciones en total. Se puede observar también que el año 2021 fue el más importante para el repositorio IEEE Xplore en el que se aportó con 15 publicaciones, el año 2022 fue crucial para ACM y Springer con 6 y 7 publicaciones respectivamente, mientras que 2023 es el año en el que ARXIV tuvo más publicaciones con un total de 6 y Science Direct mantiene un promedio de 2 publicaciones en los años 2022 a 2024.

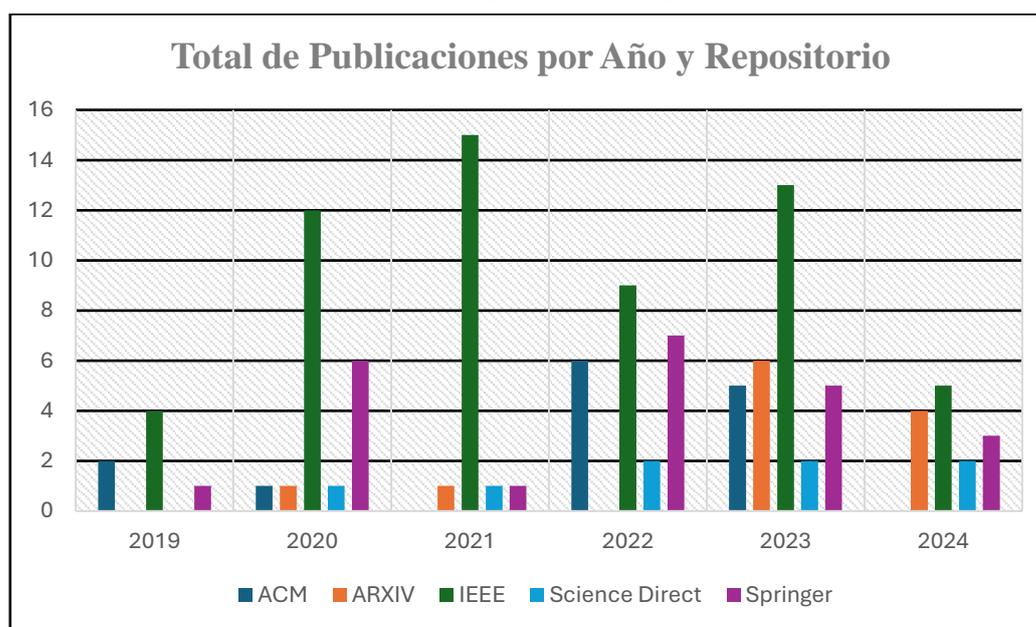


Fig. 8: Diagrama de barras total de publicaciones por año y repositorio.

II.2.3. Análisis por región geográfica

El análisis por región geográfica de las publicaciones realizadas en los repositorios es importante porque nos brinda una visión del panorama científico global, nos permite conocer diversidad de perspectivas, identificar colaboraciones internacionales, el impacto que está teniendo la investigación de este campo en ciertos países y la distribución de este contenido, así como la evolución de las temáticas y patrones de investigación. En la Figura 9 se muestran los mapas de proyección de cada uno de los repositorios, así como el volumen de publicaciones en cada una de las regiones. Para ACM podemos observar que China se sitúa en primer lugar con 11 publicaciones; en ARXIV Estados

Unidos y China están primeros con 4 publicaciones cada uno; para IEEE Xplore China es el país que más aporta con 16 publicaciones seguido de Estados Unidos con 7; Science Direct sitúa también a China en primer lugar con 7 publicaciones al igual que el repositorio Springer con 9.

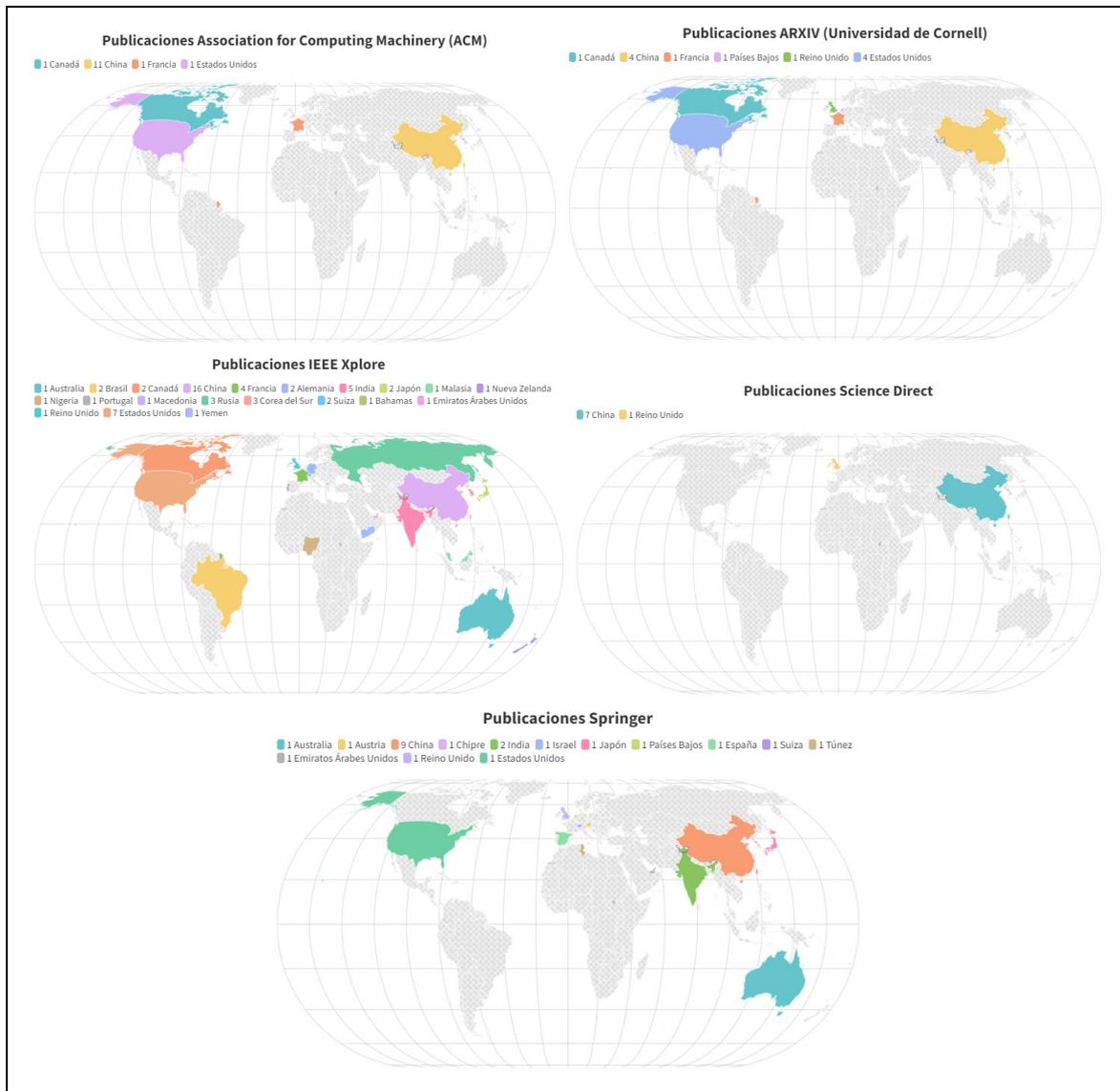


Fig. 9: Mapa de proyección ACM/ARXIV/IEEE/Science Direct/Springer.

A nivel global la Figura 10 representa el mapa de proyección de todas las regiones geográficas que aportan con artículos científicos en el área de codificación de vídeo con IA. China está situada en primer lugar con un total de 47 aportaciones que puede deberse a la combinación de varios factores como la inversión masiva en investigación y desarrollo, exploración e innovación tecnológica constante, colaboración entre la academia e industria, accesibilidad a grandes volúmenes de datos, información y potente computación. Seguido tenemos a Estados Unidos con 13 publicaciones, la India con 7, Francia con 6, Canadá y Reino Unido con 4. El continente Asiático, Europeo y América del Norte aportan con la mayor cantidad de publicaciones para este campo de investigación, producto de la combinación de factores económicos, educativos y tecnológicos que contribuyen con la investigación, evolución y difusión de los nuevos estándares y modelos de codificación.

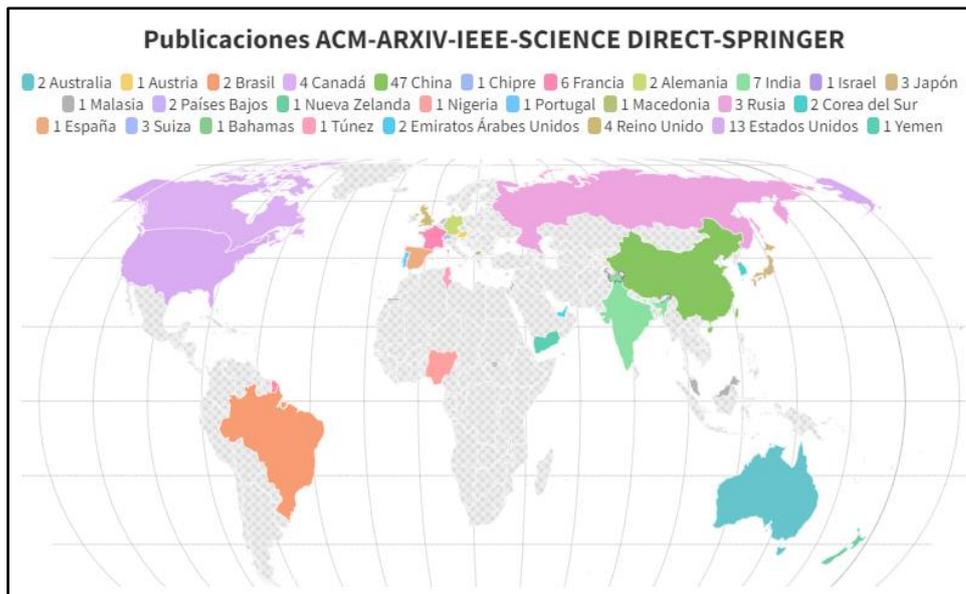


Fig. 10: Mapa de proyección publicaciones total de repositorios.

II.2.4. Análisis bibliométrico por autor

El análisis bibliométrico por autor nos permite identificar y visualizar las relaciones de colaboración, autores más influyentes y comprender las redes de coautoría. VOSviewer es la herramienta que nos permite visualizar estas características a través de estructuras en forma de red compuestas de clústeres, nodos y conexiones. Cada autor es representado por un nodo y las conexiones entre nodos representan las colaboraciones en publicaciones o coautorías en un artículo, el número de conexiones que tiene un autor con otros nodos indica su nivel de colaboración y las conexiones indican la densidad de colaboración entre autores. El tamaño de los nodos representa la cantidad de publicaciones de un autor, por lo que un nodo de mayor tamaño indica su nivel de actividad y contribución. Los clústeres representan un grupo de autores que colaboran entre sí en una investigación por lo que son de utilidad a la hora de identificar las redes de colaboración. La centralidad de los autores en la red puede darnos una idea de su influencia dentro de una publicación, por lo que los nodos centrales pueden ser considerados como autores principales o importantes.

La Figura 11 muestra las redes bibliométricas por autor de cada repositorio analizado, los artículos cuyos autores no tienen relaciones de colaboración con otras publicaciones no se muestran en la red. La red ACM se compone de tres clústeres con los artículos [8], [10] y [7] los autores más influyentes son Zhao Tiesong y Liu Jiaying; por otro lado, ARXIV está formado por los clústeres [18], [24] y [19] donde se puede observar una importante influencia de los autores Zhao Yanchen, He Wenxuan y Wang Qizhe ubicados centralmente en la red. IEEE está compuesto por los clústeres [30], [31] y [80], de los cuales He Gang, Wu Chang, Li Lei y Xie Weiyang destacan como los más relevantes. Los repositorios Science Direct y Springer tienen dos clústeres en sus redes, el primero destaca por las publicaciones [90] y [87] con los autores Kwong Sam y Wang Shiqi; por último, Springer tiene [113] y [99] como artículos destacados con Jia Chuanmin y Ma Siwei como sus autores influyentes.

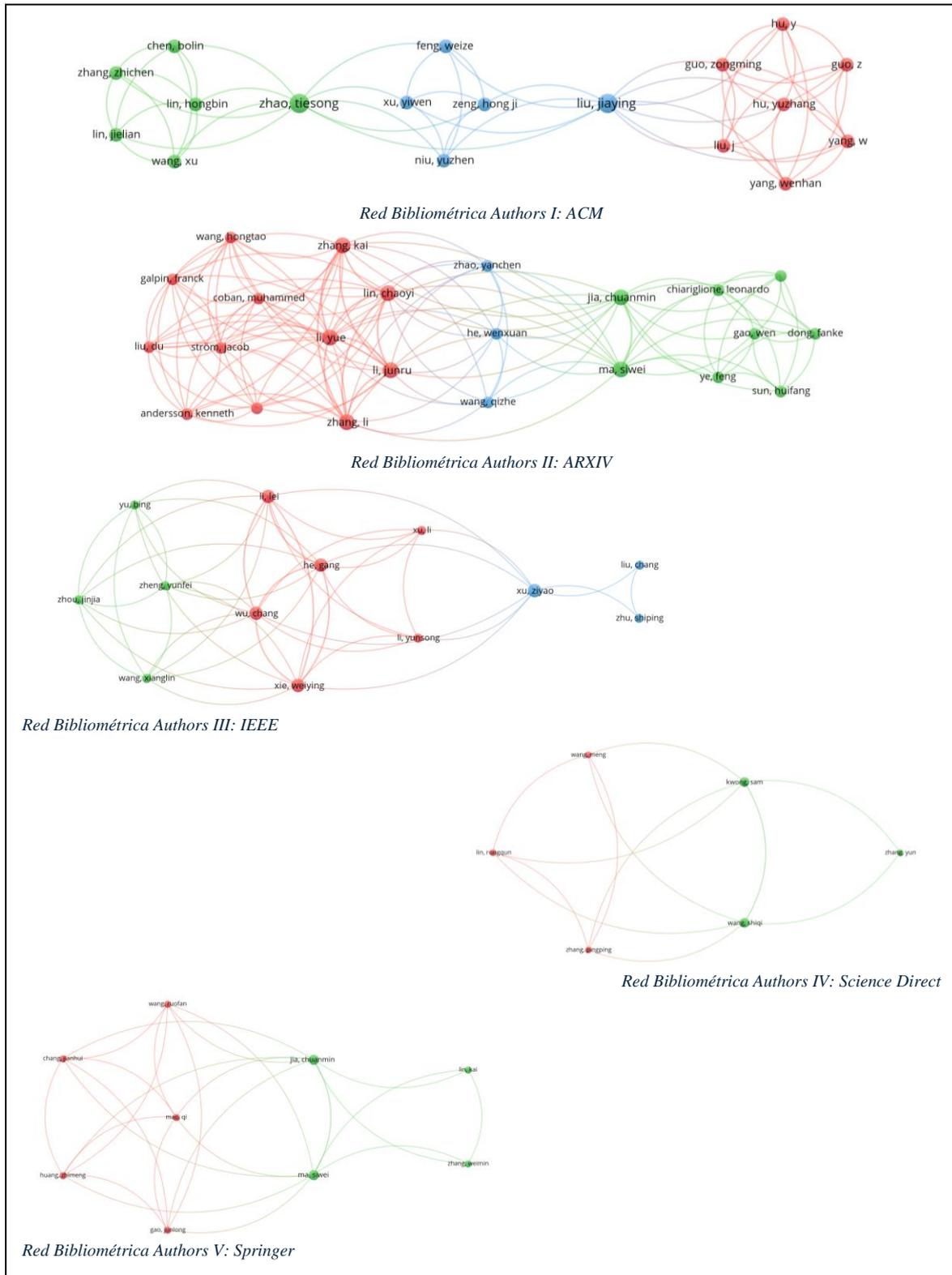


Fig. 11: Red bibliométrica por autor ACM/ARXIV/IEEE/Science Direct/Springer.

La estructura global de la red bibliométrica trazada por todos los autores de las publicaciones presentados en todos los repositorios se muestra en la Figura 12, es una red formada por seis clústeres en los cuales se destacan los nodos centrales correspondientes a Jia Chuanmin y Ma Siwei que son los autores centrales de la red y por lo tanto considerados los más influyentes.

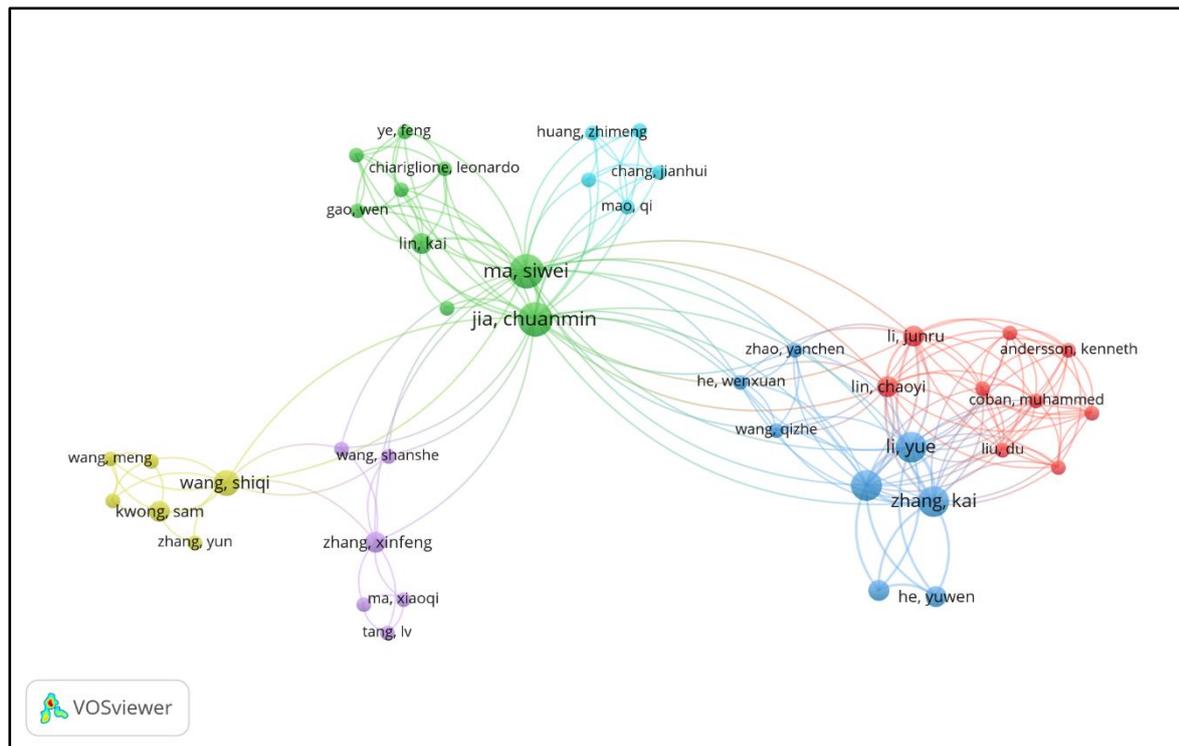


Fig. 12: Red bibliométrica por autor publicaciones total de repositorios.

II.2.5. Análisis bibliométrico por palabras clave

El análisis bibliométrico por palabras clave nos permite identificar y visualizar las tendencias de investigación, los temas emergentes y las áreas de interés con más investigación. VOSviewer es la herramienta que nos permite visualizar los patrones y tendencias de investigación a través de la evaluación de la frecuencia y concurrencia de palabras clave. En la estructura de esta red los nodos están representados por palabras clave, su tamaño indica la frecuencia de uso dentro de los artículos. Las conexiones entre nodos nos muestran las concurrencias de las palabras clave dentro de un artículo y una distancia cercana entre ellos refleja que las palabras tienen una relación temática fuerte. Los clústeres por su parte engloban los posibles subtemas dentro de un campo de estudio y están representados por colores dentro de la red.

La Figura 13 muestra las redes bibliométricas por palabras clave de cada repositorio analizado, aquellas palabras clave que no tienen ninguna relación de concurrencia en las publicaciones no se visualizan en la estructura de la red. La red ACM se compone de ocho clústeres donde destacan las palabras *Video Compression*, *Deep Learning* y *Video Coding* por ser los nodos de mayor tamaño y con mayor número de conexiones indicando su alta frecuencia de uso dentro de los artículos. Por otro lado, el repositorio ARXIV está formado por un solo clúster con ocho palabras clave que guardan una concurrencia temática entre sí. IEEE es una red con una estructura bastante compleja formada por varios nodos, clústeres y conexiones, en total se contabilizan aproximadamente 15 clústeres donde resaltan las palabras *Video Compression*, *Deep Learning*, *Video Coding* y *Learned Video Compression*. La red Science Direct está formada por tres clústeres fácilmente identificables y de

aspecto simétrico, las palabras *Deep Learning* y *Video Coding* son las protagonistas de la red por su centralidad y número de conexiones. Por último, Springer es una red con nueve clústeres que resaltan a las palabras *Deep Learning*, *Video Compression*, *Video Coding* y *Video Super-Resolution*.

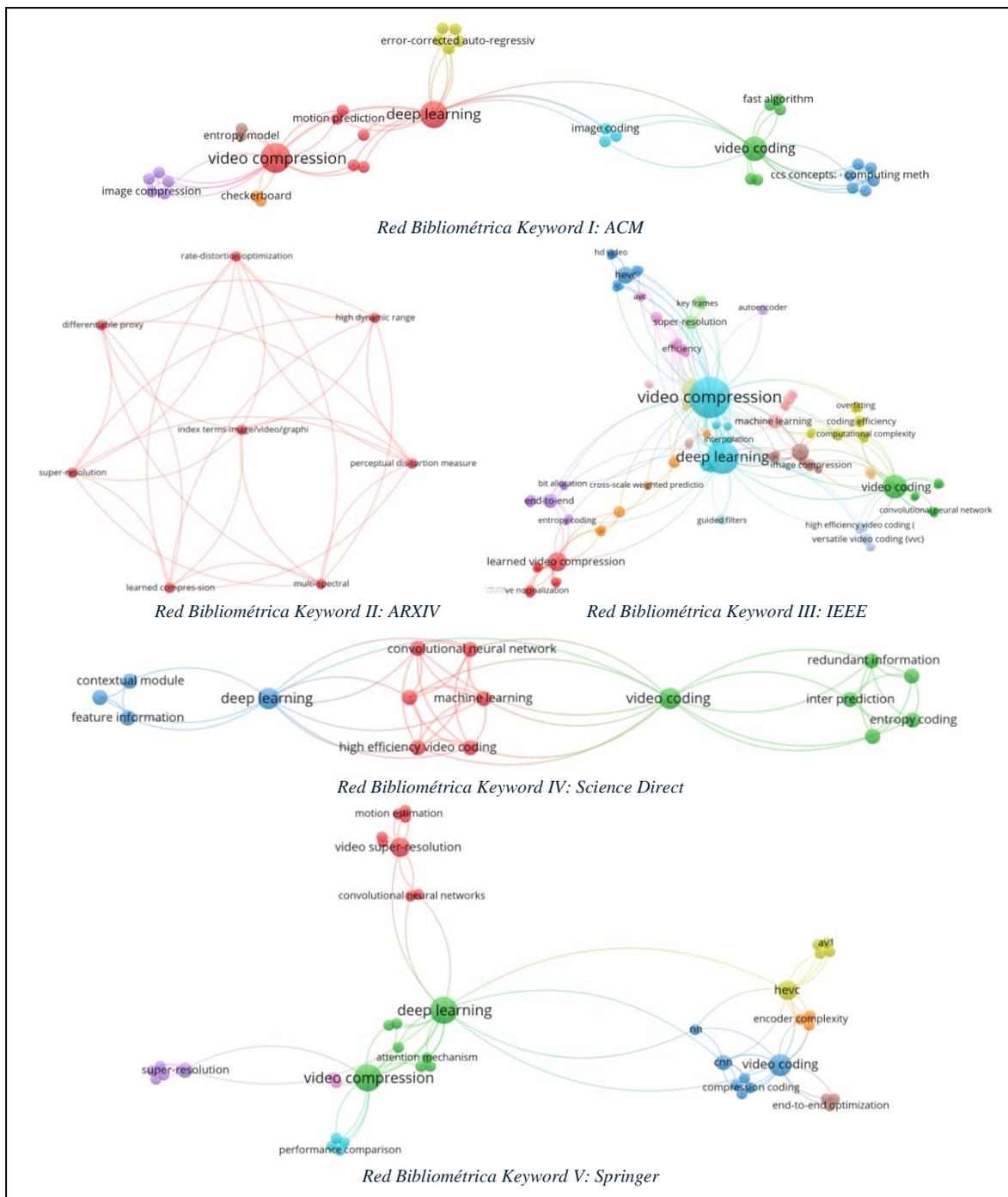


Fig. 13: Red bibliométrica por palabras clave ACM/ARXIV/IEEE/Science Direct/Springer.

La Figura 14 muestra la estructura global de la red formada por todas las palabras clave del total de repositorios analizados. Como se observa es una red bastante amplia y compleja con una gran cantidad de nodos, clústeres y conexiones, sin embargo, es posible visualizar cuáles con las palabras clave más importantes. En primer lugar, *Video Compression*, seguida de *Video Coding* y por último *Learned Video Compression* consideradas como las de mayor uso frecuente dentro de los artículos.

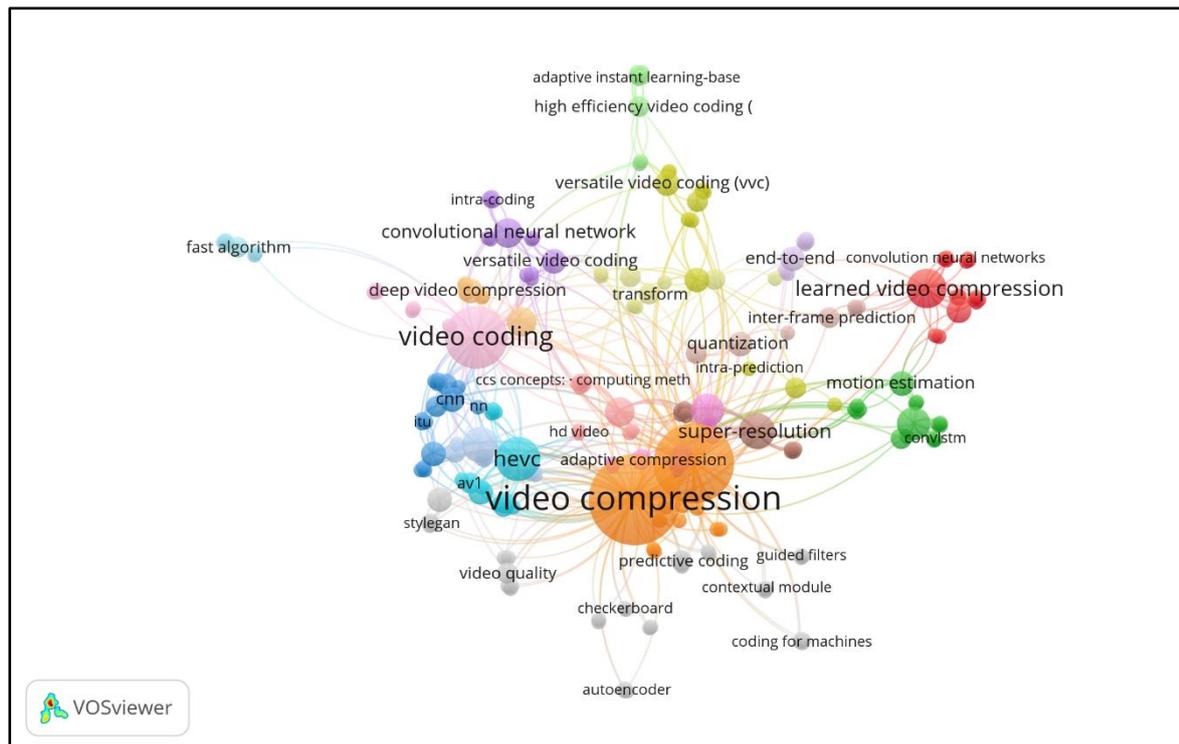


Fig. 14: Red bibliométrica por palabras clave publicaciones total de repositorios.

II.2.6. Análisis bibliométrico por título y resumen

El análisis bibliométrico por título y resumen nos permite examinar la manera en que los términos y conceptos se agrupan y relacionan entre sí dentro de las publicaciones, el título y resumen son claves en la estructura de un artículo debido a que nos indican la temática central, el área de estudio y nos brinda la idea principal del contenido de la publicación. VOSviewer es la herramienta que nos permite construir las redes que muestran cómo los términos importantes se agrupan y sugieren los enfoques o tendencias de investigación. En la estructura de esta red los términos que aparecen frecuentemente juntos en títulos y resúmenes forman clústeres que representan temáticas de investigación relacionadas. Cada término es representado por un nodo y su tamaño nos puede dar una idea de la relevancia dentro de un campo de estudio o de su frecuencia de uso dentro de las publicaciones. Las conexiones entre los nodos nos indican las concurrencias de los términos y muestran las relaciones fuertes que existen entre ellos.

La Figura 15 muestra las redes bibliométricas por título y resumen de cada repositorio analizado, aquellos términos que no guardan ninguna relación de concurrencia en las publicaciones son omitidos en la estructura de la red. La red ACM está formada por varios nodos de donde sobresalen los términos *Video Coding*, *Vídeo*, *Entropy Model*, *Algorithm* y *Frame*. ARXIV por su parte tiene a los términos *Neuronal Network*, *Model*, *Vídeo*, *Deep Learning* y *Artificial Intelligence* como centrales de mayor uso frecuente. La red IEEE es una estructura muy compleja y compacta se puede observar que el término central de la red es *Video Compression*, por detrás aparecen los términos *Deep Learning*, *Image*, *Learning Model*, *Study* y *Artificial Intelligence* que guardan estrecha relación en

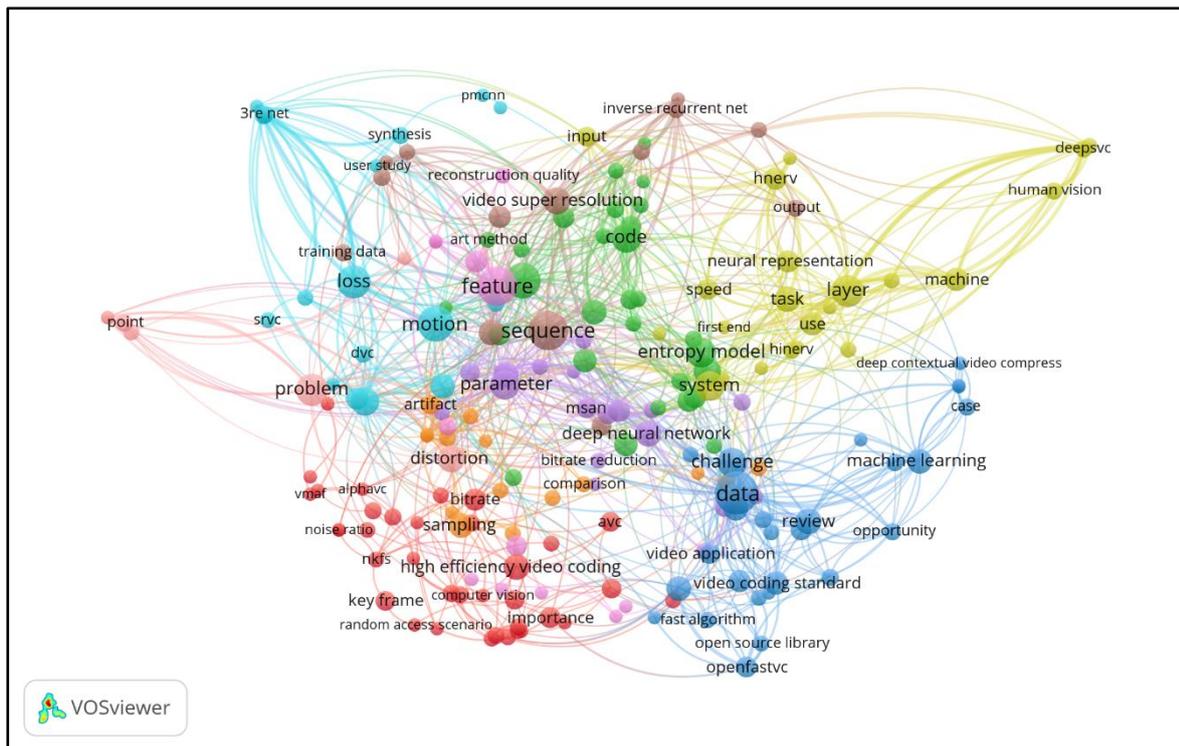


Fig. 16: Red bibliométrica por título y resumen publicaciones total de repositorios.

II.3. ANÁLISIS CUALITATIVO

La sección dedicada al análisis cualitativo tiene como objetivo analizar el estado actual de la evolución de la codificación de vídeo, el enfoque tradicional, el nuevo enfoque IA, las métricas para evaluación de prestaciones y las publicaciones que brindan acceso a código abierto en GitHub para probar nuevos modelos de codificación de vídeo con IA, proporcionando una comprensión global para entender la tecnología actual, desafíos y desarrollo futuro.

II.3.1. Evolución de la codificación de vídeo

La evolución de la codificación de vídeo tiene como protagonistas a los estándares de codificación y a los códecs asociados a ellos que son las herramientas que permiten la compresión de vídeo. Estos aplican algoritmos que buscan eliminar redundancias, reduciendo el tamaño del archivo o la velocidad de bits si es una transmisión en vivo, optimizando los requisitos de ancho de banda y disponibilidad de almacenamiento [35].

En el año de 1984 aparece H.120, el primer estándar de códec de vídeo que aplicaba técnicas de conmutación, cuantificación y modulación diferencial; para 1990 se introduce H.261 desarrollado para videoconferencias empleando una codificación híbrida basada en bloques; en 1993 aparece MPEG-1 implementado ampliamente en los CD de vídeo. Para el año de 1995 se terminó H.262/MPEG-2 empleado ampliamente en los DVD y la televisión digital y en el periodo de 1996 a 1998, H.263/H.263+ fue desarrollado logrando una compresión de vídeo de baja tasa de bits y aplicado para vídeo a través de conexión inalámbrica 3G y videoconferencia telefónica [27].

En 1999 la codificación de vídeo avanza con MPEG-4 que se completa en el año 2003 coeditado con H.264/AVC que soporta vídeo en internet, computadores, transmisión inalámbrica y redes móviles, televisión HD, Blue-Ray, HD DVD, cable y satélite, ofrece mejor calidad con tasas de bits más bajas y compensación de movimiento mejorada. Luego de varios años de desarrollo e investigación en 2013 aparece el estándar de códec de vídeo H.265 o de alta eficiencia desarrollado para vídeo de Ultra HD de hasta 8k con hasta 1230 fps y un mayor ahorro de velocidad de bits en comparación con su predecesor. Paralelamente entre los años de 2010 a 2015 se presentan los códecs estandarizados VP8, VP9 y VP10 desarrollados por Google y de libres regalías [13].

AV1 fue presentado en el año 2018 y desarrollado por *Alliance for Open Media* como una tecnología multimedia abierta y libre de regalías, es usado en la transmisión de vídeo en línea para plataformas de streaming como YouTube, Netflix, Amazon Prime Video y Vimeo [66]. El último estándar de códec de vídeo presentado en 2021 es H.266 o de codificación de vídeo versátil (VVC) planteado para la aplicación de experiencia inmersiva, transmisión de realidad virtual y vídeo de alto rango dinámico HDR [35]. La cronología completa sobre la evolución de la codificación de vídeo se ilustra en la Figura 17 que presenta los estándares de códecs de vídeo, así como las organizaciones y organismos de estandarización que trabajaron en cada uno de ellos.

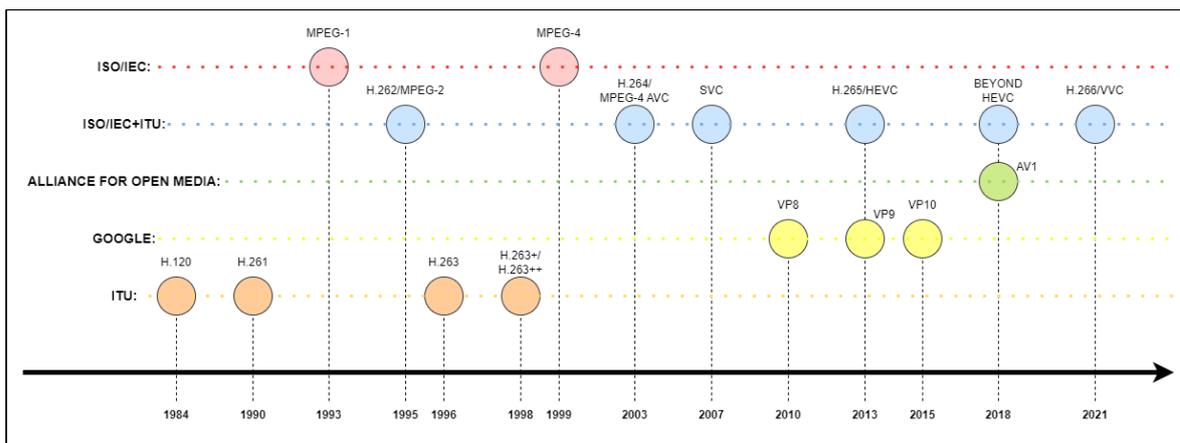


Fig. 17: Diagrama línea de tiempo evolución codificación de vídeo.

II.3.2. Enfoque tradicional para la codificación de vídeo

La codificación de vídeo se centra en dos estructuras principales, el codificador y el decodificador. El codificador de vídeo tiene tres componentes o módulos: temporal, espacial y codificación entrópica [27]. La codificación de vídeo busca y elimina redundancias que pueden ser de cuatro tipos: espacial, temporal, estadística y espacio de color, y los módulos del codificador aprovechan estas redundancias para reducir la cantidad de datos logrando que el archivo multimedia reduzca su tamaño, pueda ser transmitido en directo y que pueda ser decodificado [35]. La compresión de vídeo puede ser de dos tipos: con pérdida o sin pérdida. La primera reduce los datos redundantes sin comprometer la calidad del vídeo, pero ofrece una reducción limitada del archivo multimedia o de la

velocidad de bits y es de carácter reversible. La segunda por otro lado ofrece una mayor reducción de datos, pero solo recupera una aproximación del vídeo de origen [35][66].

Un vídeo se compone de una secuencia de fotogramas que se comprimen según su tipo y aprovechando su redundancia y estos fotogramas pueden ser del tipo I o trama *Intra*, P o trama *Predictive* y B o trama *Bidirectional*. El vídeo se puede comprimir por dos etapas, *intraframe* e *interframe*. La primera comprime el fotograma de forma individual o autónoma respecto de los demás fotogramas y la segunda aprovecha los datos redundantes entre fotogramas sucesivos [27][65]. La codificación de vídeo de enfoque tradicional y sus estándares se valen de las técnicas de compresión entre cuadros, predicción temporal, codificación de transformación, cuantización, codificación entrópica, compensación y estimación de movimiento [27], [35], [66].

Los estándares de códecs de vídeo AV1, H.264/AVC, VP9, H.265/HEVC y el actual H.266 (VVC) utilizan este enfoque. H.264 o AVC utiliza el enfoque de estimación de movimiento basado en bloques y es empleado ampliamente para aplicaciones de videovigilancia, videoconferencia, almacenamiento digital, transmisión de televisión e Internet. H.265 o HEVC con su codificación de vídeo de alta eficiencia permitió la transmisión y streaming en 4K de alto rango dinámico HDR. El estándar H.266 o VVC utiliza un enfoque híbrido basado en bloques con las tecnologías de HEVC, es un códec de vídeo versátil que utiliza herramientas y métodos como árboles de codificación y estructura de bloques, fue desarrollado para amplias aplicaciones de vídeo 8K o más, realidad inmersiva o aumentada [35], [65], [66].

II.3.3. Enfoque IA para la codificación de vídeo

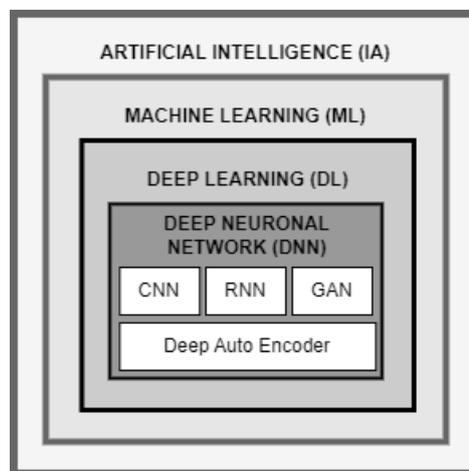


Fig. 18: Enfoque codificación vídeo con IA.

El enfoque de codificación de vídeo con Inteligencia Artificial (IA) abarca conceptos de *Machine Learning* (ML), *Deep Learning* (DL) y *Deep Neuronal Network* (DNN), los cuales se utilizan ampliamente para describir sistemas que se comportan de manera inteligente. El término raíz es inteligencia artificial que vincula el conocimiento humano con la inteligencia de máquinas, de aquí se desprende el término *machine learning* que se enfoca en los datos y experiencia construyendo modelos analíticos. Más adentro encontramos el *deep learning* que utiliza métodos de aprendizaje

utilizando redes neuronales multicapa, por tanto, el aprendizaje profundo hace referencia a múltiples niveles o etapas de aprendizaje [120].

La Figura 18 muestra el enfoque de la codificación de vídeo con IA. Su núcleo está en el aprendizaje profundo (DL) con las redes neuronales profundas (DNN), dentro de su clasificación en primer lugar encontramos las redes neuronales convolucionales (CNN) que consta de varias capas sucesivas que pueden entrenarse para tareas de predicción [113]. Luego encontramos a las redes neuronales recurrentes (RNN) que consideran las etapas pasadas para etapas futuras por lo que poseen propiedades de memoria adecuada para tratar datos de carácter secuencial [76] y por otro lado tenemos a las redes generativas antagónicas (GAN) que son utilizadas ampliamente para crear muestras o aprender patrones de datos utilizados posteriormente para generar nuevos datos con propiedades similares al original [120]. Por último, encontramos al codificador automático profundo que utiliza redes neuronales para aprender a representar datos reduciendo su dimensionalidad [76], y en conjunto todas estas técnicas resultan de utilidad para cumplir los objetivos de la compresión de vídeo.

II.3.4. Métricas para evaluación de prestaciones en la codificación de vídeo

Las métricas de evaluación en la codificación de vídeo son herramientas que nos ofrecen una medida de la calidad, eficiencia y rendimiento de los algoritmos de compresión de vídeo. Existen dos formas de evaluar la calidad de un contenido multimedia, las métricas objetivas y las métricas subjetivas. La diferencia principal entre ambas radica en que las primeras analizan parámetros técnicos de una señal a través de cálculos numéricos y las segundas pretenden obtener una medida equivalente a la percepción visual o experiencia del consumidor.

Peak Signal Noise Ratio (PSNR) es una métrica de calidad objetiva que se utiliza para medir la calidad de reconstrucción de vídeo después de la compresión y descompresión [92], se calcula utilizando el error cuadrático medio (MSE) y su resultado se expresa en decibeles (dB). La expresión matemática para calcular el PSNR del vídeo original X y el vídeo distorsionado Y se muestra en la Ecuación 1.

$$PSNR(X, Y) = 10 \log_{10} \left(\frac{255^2}{MSE(X, Y)} \right) \quad [dB] \quad (1)$$

Mean Opinion Score (MOS) es una métrica de calidad subjetiva que tiene por objetivo obtener una medida de la percepción del usuario frente a un contenido multimedia con una escala de evaluación entre 1 a 5, y por lo general se aplica mediante encuestas y consiste en puntuar una serie de vídeos con diferentes tasas de codificación para calificar la calidad de los vídeos [121].

Video Multimethod Assessment Fusion (VMAF) es una métrica de evaluación de calidad de vídeo perceptual desarrollada por Netflix y que considera para su evaluación factores visuales a través del aprendizaje automático y conjuntos de datos subjetivos por lo que es un algoritmo que combina la calidad objetiva y subjetiva [122].

II.3.5. Publicaciones con código abierto en GitHub para la codificación de vídeo

GitHub es una plataforma destinada al desarrollo colaborativo de software que permite a equipos y desarrolladores gestionar y almacenar código fuente de software, brindándonos la posibilidad de acceder a proyectos de código abierto que cuentan con amplios recursos, herramientas y conocimiento facilitando el aprendizaje, colaboración, innovación y solución a problemas.

El crecimiento del campo de la codificación de vídeo con IA en gran medida se debe al desarrollo de nuevos algoritmos de compresión. Actualmente se investigan diferentes enfoques para desarrollar estos algoritmos partiendo de machine learning, aprendizaje profundo y redes neuronales y sobre estos algoritmos se realizan pruebas para evaluar sus prestaciones como rendimiento y eficiencia. Para este trabajo se han clasificado las publicaciones que cuentan con repositorios de código abierto en GitHub y se resumen en la tabla 2.

REPOSITORIO	PUBLICACIONES	AÑO
ACM	[4], [8]	2022, 2023
ARXIV	[15], [16], [26]	2020, 2023, 2023
IEEE	[37], [75], [77], [78], [79]	2023, 2022, 2021, 2023, 2023
SCIENCE DIRECT	[85], [86]	2022, 2022
SPRINGER	[107], [110]	2020, 2020

Tabla 2: Publicaciones código abierto GitHub.

El número total de publicaciones consideradas para este estudio es 115 tomadas de los repositorios ACM, ARXIV, IEEE Xplore, Science Direct y Springer, y como se observa en la Tabla 2 existen un número muy reducido de publicaciones que cuentan con código abierto en la plataforma de GitHub, en total 14, e IEEE es la que más aportes realiza. Una vez realizado el análisis para examinar cada uno de los repositorios disponibles con código y tomando en cuenta factores importantes como, año más reciente de publicación, disponibilidad de modelos de entrenamiento actualizados, instrucciones de implementación y ejecución específicas, acceso a ficheros completos de instalación y disponibilidad actual del repositorio, se han seleccionado dos soluciones propuestas del repositorio más influyente de este estudio IEEE Xplore [77] y [75] para ser implementadas y evaluar sus prestaciones frente a los estándares de codificación tradicional como son H.264 y H.265.

Learning Cross-Scale Weighted Prediction for Efficient Neural Video Compression (ENVC), propone un modelo de códec neuronal con un módulo de predicción de escala cruzada para lograr una compensación de movimiento más efectiva, además de incorporar una estrategia de cuantificación de múltiples etapas [77]. Este modelo permite ajustar el tamaño del GoP trabajando con frames de tipo I y P con valores de Bitrate de 512, 1024, 2048, 3072 y 4096 kbps. *Advancing Learned Video Compression With In-Loop Frame Prediction* (ALVC), propone un modelo de compresión de aprendizaje avanzado que incorpora un módulo de predicción de cuadros en bucle con la capacidad de predecir los frames siguientes a partir de los comprimidos previamente [75], esta propuesta trabaja con frames de tipo I, P y B, permite ajustar el tamaño del GoP y trabaja con valores de Bitrate de 256, 512, 1024 y 2048 kbps.

III.EVALUACIÓN DE PRESTACIONES

III.1.ANÁLISIS PRE-CODIFICACIÓN

El análisis de pre-codificación tiene por objetivo mostrar las características de los vídeos que se utilizaran para evaluar las prestaciones de las soluciones de codificación de vídeo con IA (ENVC y ALVC). Se analizan cuatro vídeos estandarizados en términos de resolución espacial y temporal, con una duración similar, mismo formato y codificación como se muestra en la Tabla 3.

III.1.1.Información general banco de vídeos de prueba

PARÁMETRO	SINTEL	FOOTBALL	SIMULATOR	ELEKTRA
DURACIÓN	00:00:09.96	00:00:10.13	00:00:10.08	00:00:10.08
CODIFICACIÓN DEL VÍDEO	h264	h264	h264	h264
FORMATO ORIGINAL (YUV)	yuv420p	yuv420p	yuv420p	yuv420p
RESOLUCIÓN	416x240	416x240	416x240	416x240
FPS	24 fps	24 fps	24 fps	24 fps
BITRATE VÍDEO	8067 kb/s	10201 kb/s	7326 kb/s	7449 kb/s

Tabla 3: Información general vídeos de prueba.

III.1.2.Información de la complejidad espacial y temporal (SITI)

INFORMACIÓN SITI	SINTEL	FOOTBALL	SIMULATOR	ELEKTRA
AVG_SI	84.020882	96.164452	136.032410	61.643539
AVG_TI	14.093035	24.607841	15.402292	7.498899

Tabla 4: Información complejidad espacial y temporal (SITI) vídeos de prueba.

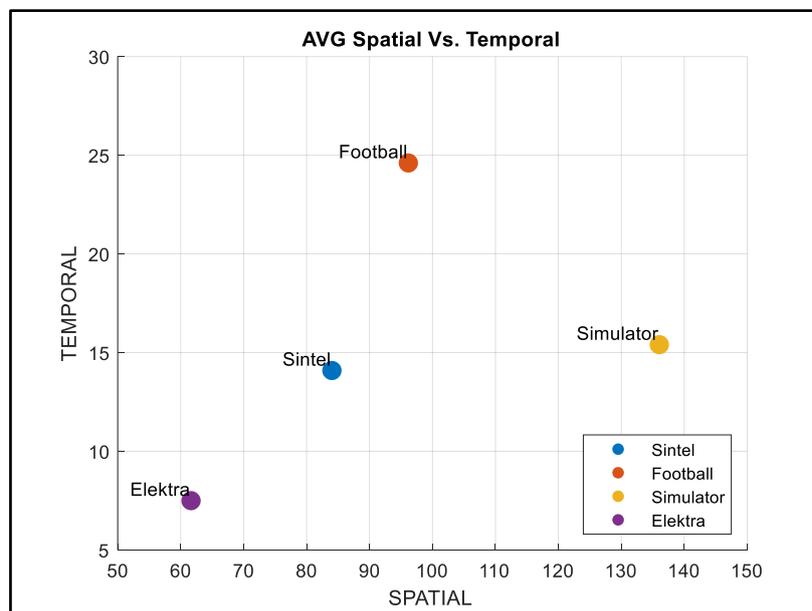


Fig. 19: Gráfica complejidad espacial y temporal (SITI) vídeos de prueba.

III.2.ANÁLISIS POST-CODIFICACIÓN

El análisis post-codificación tiene por objetivo realizar la evaluación de la calidad visual a través de la representación gráfica de la métrica de calidad objetiva PSNR y la métrica de calidad perceptual VMAF de cada uno de los vídeos de prueba codificados con H.264, H.265, ENVC y ALVC a diferentes Bitrates y con diferentes tamaños de GoP.

III.2.1. Valores PSNR GoP 12 codificación frames tipo IP

PSNR - GROUP OF PICTURES GOP 12 (IP)									
VÍDEO	CODIFICACIÓN TRADICIONAL					CODIFICACIÓN IA			
	Bitrate [kbps]	Bitrate Real [kbps]	PSNR H.264	Bitrate Real [kbps]	PSNR H.265	Bitrate Real [kbps]	PSNR ENVC	Bitrate Real [kbps]	PSNR ALVC
SINTEL	4096	3969	48.4257	3862	48.9631	810.165	37.4268		
	3072	2984	46.0734	2909	46.4778	739.421	36.9891		
	2048	2010	43.2929	1936	43.6593	549.132	35.4650	628.121	36.0135
	1024	1014	39.4442	971	39.9783	361.570	33.8523	378.536	33.8689
	512	510	36.1320	484	36.4269	230.424	32.1108	243.038	31.7258
	256	256	33.0335	241	33.1086			147.766	29.3873
FOOTBALL	4096	4085	46.0786	4032	46.4993	1117.650	35.9223		
	3072	3063	44.0353	3041	44.4598	1008.479	35.5056		
	2048	2047	41.5271	2033	41.9676	769.826	34.1199	963.735	33.9657
	1024	1027	37.8626	1011	38.3917	532.064	32.7379	568.129	31.8158
	512	514	34.5880	507	35.2697	356.749	31.2036	362.093	29.9338
	256	253	31.5107	254	32.3854			220.925	28.1304
SIMULATOR	4096	4096	50.1156	4126	49.8221	941.916	36.7309		
	3072	3109	47.7641	3094	47.4214	860.901	36.3711		
	2048	2078	44.6094	2062	44.4146	652.929	34.8258	730.349	34.9736
	1024	1039	40.0181	1032	40.0419	442.110	33.2881	444.966	32.5022
	512	520	36.2354	516	36.1340	291.097	31.6058	285.622	30.4820
	256	260	32.6790	258	32.9443			177.252	28.2027
ELEKTRA	4096	4166	51.8668	4125	52.3609	266.546	40.4165		
	3072	3123	50.4940	3093	51.0199	244.640	40.0999		
	2048	2083	48.9731	2062	49.5879	202.559	39.2370	268.842	39.5480
	1024	1035	46.8779	948	47.1678	140.304	37.8373	170.802	37.4779
	512	515	44.6893	447	44.6035	95.726	36.2732	116.987	35.8179
	256	253	41.9205	215	41.5084			82.633	33.7020

Tabla 5: Valores PSNR GoP 12 codificación frames tipo IP.

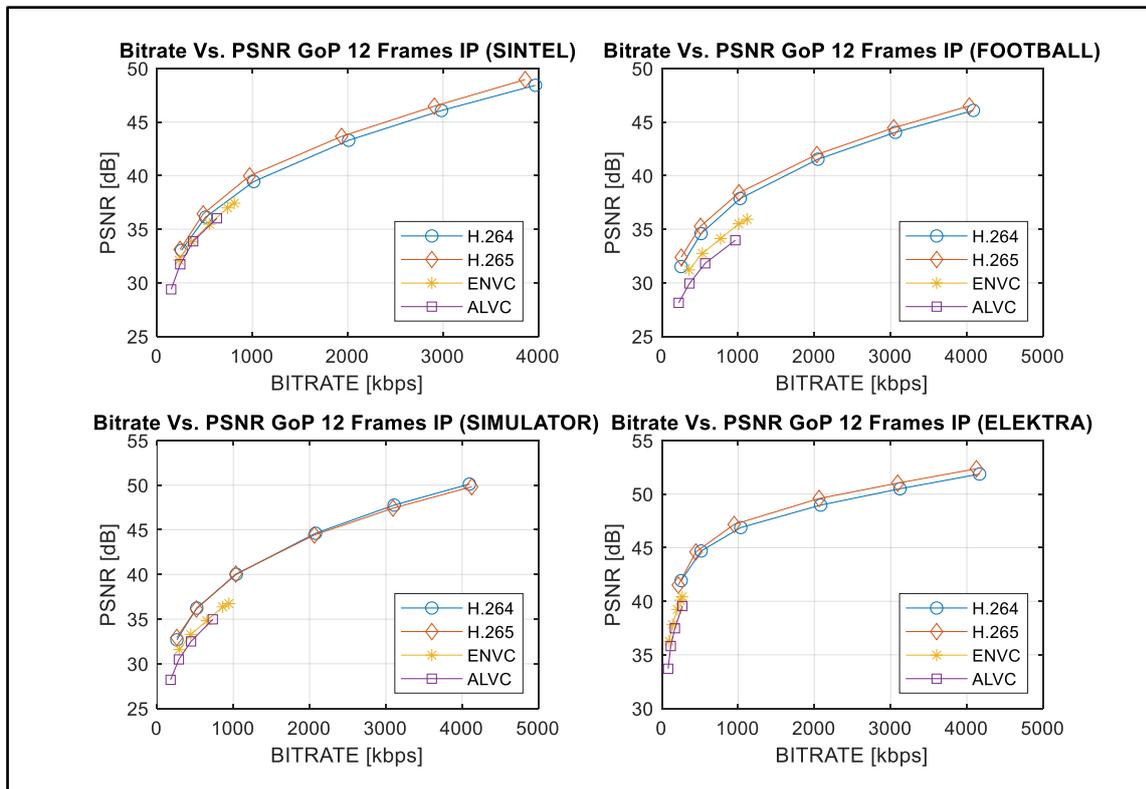


Fig. 20: Gráficas PSNR GoP 12 frames IP.

III.2.2. Valores PSNR GoP 24 codificación frames tipo IP

PSNR - GROUP OF PICTURES GOP 24 (IP)									
VÍDEO	CODIFICACIÓN TRADICIONAL					CODIFICACIÓN IA			
	Bitrate [kbps]	Bitrate Real [kbps]	PSNR H.264	Bitrate Real [kbps]	PSNR H.265	Bitrate Real [kbps]	PSNR ENVC	Bitrate Real [kbps]	PSNR ALVC
SINTEL	4096	3956	49.0052	3729	49.0585	696.803	36.8640		
	3072	2981	46.6629	2822	46.8586	639.796	36.5389		
	2048	2007	43.8124	1876	44.0470	462.281	34.9138	509.853	35.0666
	1024	1015	39.9044	930	40.1550	286.077	33.1422	312.012	32.9405
	512	510	36.7340	459	36.6330	174.186	31.3755	223.595	31.0510
	256	251	33.6055	223	33.3442			138.280	28.9343
FOOTBALL	4096	4053	46.4308	3900	46.6392	1012.926	35.4572		
	3072	3040	44.4381	2931	44.6238	920.863	35.1131		
	2048	2031	41.9830	1952	42.1167	689.286	33.6268	829.310	33.2091
	1024	1003	38.2919	965	38.5250	470.440	32.2117	496.963	31.1227
	512	499	35.0283	478	35.3769	313.051	30.6807	330.430	29.3449
	256	248	31.9435	237	32.5458			205.350	27.6204
SIMULATOR	4096	4020	50.3258	4013	49.7792	843.247	36.2051		
	3072	3043	48.0245	3010	47.4368	775.052	35.9111		
	2048	2052	45.0207	2017	44.5045	575.830	34.2485	610.110	34.1732
	1024	1034	40.5541	1020	40.2798	376.604	32.6125	368.347	31.7569
	512	517	36.9041	515	36.6416	239.515	30.9023	249.572	29.7982
	256	257	33.4711	258	33.4465			154.988	27.7117
ELEKTRA	4096	4162	52.0461	4124	52.4970	216.136	40.0633		
	3072	3124	50.6922	3093	51.1665	203.537	39.8280		
	2048	2081	49.2148	2062	49.7459	163.827	38.8810	223.600	39.1170
	1024	1036	47.2104	1031	47.3101	110.374	37.3964	148.215	37.0019
	512	510	45.1785	511	45.5954	74.072	35.7492	110.968	35.4477
	256	250	42.6239	252	43.3720			75.186	33.3887

Tabla 6: Valores PSNR GoP 24 codificación frames tipo IP.

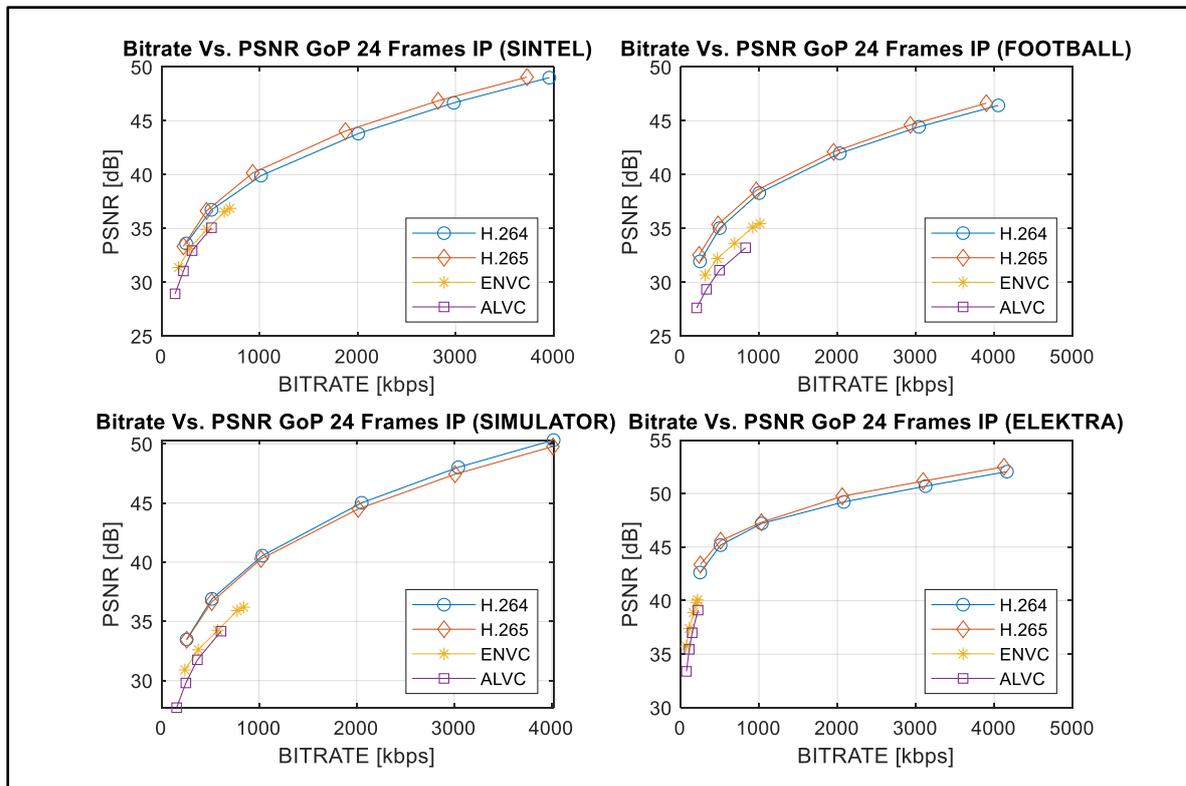


Fig. 21: Gráficas PSNR GoP 24 frames IP.

III.2.3. Valores PSNR GoP 12 codificación frames tipo IPB

PSNR - GROUP OF PICTURES GOP 12 (IPB)							
VÍDEO	BITRATE [kbps]	CODIFICACIÓN TRADICIONAL			CODIFICACIÓN IA		
		Bitrate Real [kbps]	PSNR H.264	Bitrate Real [kbps]	PSNR H.265	Bitrate Real [kbps]	PSNR ALVC
SINTEL	2048	1880	44.2797	1966	45.2583	601.581	36.0032
	1024	961	40.5253	1001	41.1459	357.717	33.9051
	512	489	37.0826	503	37.3392	223.367	31.7471
	256	247	33.7637	252	34.0836	131.333	29.4125
FOOTBALL	2048	1988	42.1651	2034	42.6663	971.378	33.8559
	1024	993	38.4628	1018	38.8833	563.745	31.7282
	512	495	35.0931	509	35.5689	358.466	29.8549
	256	244	31.9247	255	32.6546	212.261	28.0472
SIMULATOR	2048	2045	45.0520	2065	45.0277	711.262	34.9131
	1024	1033	40.7206	1031	40.6650	427.266	32.4683
	512	518	36.9031	515	36.9221	268.442	30.4805
	256	258	33.2491	257	33.5712	163.228	28.2215
ELEKTRA	2048	2037	48.9386	2059	49.7348	260.015	39.5648
	1024	997	47.0040	1027	47.9059	161.719	37.5154
	512	490	45.0276	513	46.0716	107.516	35.8312
	256	240	42.3384	257	43.6274	74.560	33.7136

Tabla 7: Valores PSNR GoP 12 codificación frames tipo IPB.

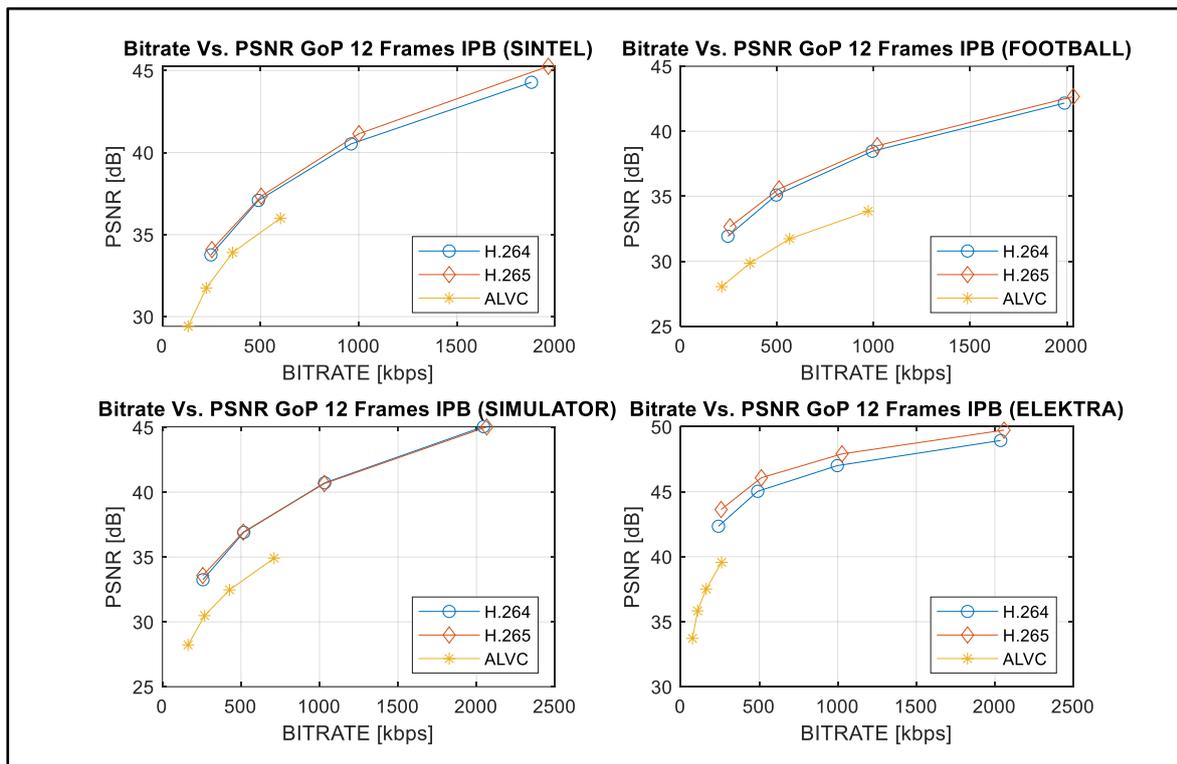


Fig. 22: Gráficas PSNR GoP 12 frames IPB.

III.2.4. Valores PSNR GoP 24 codificación frames tipo IPB

PSNR - GROUP OF PICTURES GOP 24 (IPB)							
VÍDEO	BITRATE [kbps]	CODIFICACIÓN TRADICIONAL				CODIFICACIÓN IA	
		Bitrate Real [kbps]	PSNR H.264	Bitrate Real [kbps]	PSNR H.265	Bitrate Real [kbps]	PSNR ALVC
SINTEL	2048	1846	44.4744	1905	45.4449	479.877	35.0349
	1024	927	40.8469	963	41.4809	289.583	32.9785
	512	463	37.5287	482	37.9414	201.764	31.0744
	256	229	34.2134	243	34.6568	122.843	28.9494
FOOTBALL	2048	1925	42.4277	2006	42.9513	817.325	33.1958
	1024	956	38.7792	997	39.2046	485.877	31.1200
	512	471	35.4220	498	35.8838	320.258	29.3493
	256	230	32.1812	238	32.6965	196.978	27.6153
SIMULATOR	2048	1930	45.1030	2026	45.1646	592.590	34.1619
	1024	979	40.9374	1004	40.8981	354.825	31.7486
	512	478	37.0867	507	37.3872	234.735	29.8170
	256	237	33.6267	256	34.1791	144.873	27.7263
ELEKTRA	2048	1939	48.9179	2047	49.8532	215.145	39.0963
	1024	936	47.0720	1007	48.0468	140.330	36.9833
	512	453	45.2354	492	46.2355	101.660	35.4535
	256	221	42.7885	241	43.9576	69.256	33.3727

Tabla 8: Valores PSNR GoP 24 codificación frames tipo IPB.

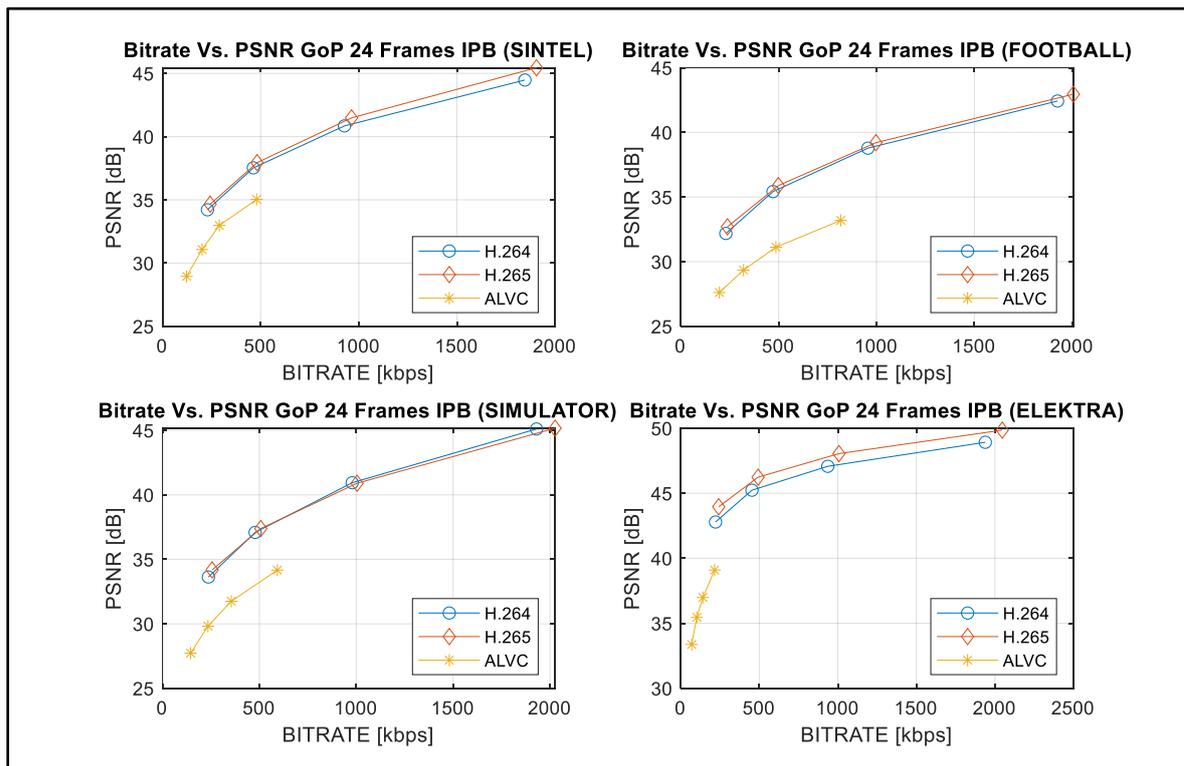


Fig. 23: Gráficas PSNR GoP 24 frames IPB.

III.2.5. Valores VMAF GoP 12 codificación frames tipo IPB

VMAF - GROUP OF PICTURES GOP 12 (IPB)							
VÍDEO	BITRATE [kbps]	CODIFICACIÓN TRADICIONAL			CODIFICACIÓN IA		
		Bitrate Real [kbps]	VMAF H.264	Bitrate Real [kbps]	VMAF H.265	Bitrate Real [kbps]	VMAF ALVC
SINTEL	2048	1880	98.7717	1966	98.9984	601.581	94.5400
	1024	961	97.2661	1001	97.6205	357.717	91.1659
	512	489	93.4298	503	93.6205	223.367	85.4383
	256	247	85.3987	252	85.9926	131.333	75.3074
FOOTBALL	2048	1988	99.3286	2034	99.4354	971.378	91.4548
	1024	993	98.1865	1018	98.0798	563.745	86.7079
	512	495	93.6462	509	93.9956	358.466	78.6320
	256	244	81.6477	255	84.6690	212.261	65.9872
SIMULATOR	2048	2045	98.7063	2065	98.6373	711.262	92.0309
	1024	1033	97.1263	1031	96.7315	427.266	89.1542
	512	518	93.2421	515	92.6929	268.442	83.6878
	256	258	84.4501	257	84.9383	163.228	73.2178
ELEKTRA	2048	2037	98.0658	2059	98.1702	260.015	93.7116
	1024	997	97.5181	1027	97.5996	161.719	90.7520
	512	490	96.3826	513	96.6157	107.516	85.9551
	256	240	92.8932	257	94.1437	74.560	78.4602

Tabla 9: Valores VMAF GoP 12 codificación frames tipo IPB.

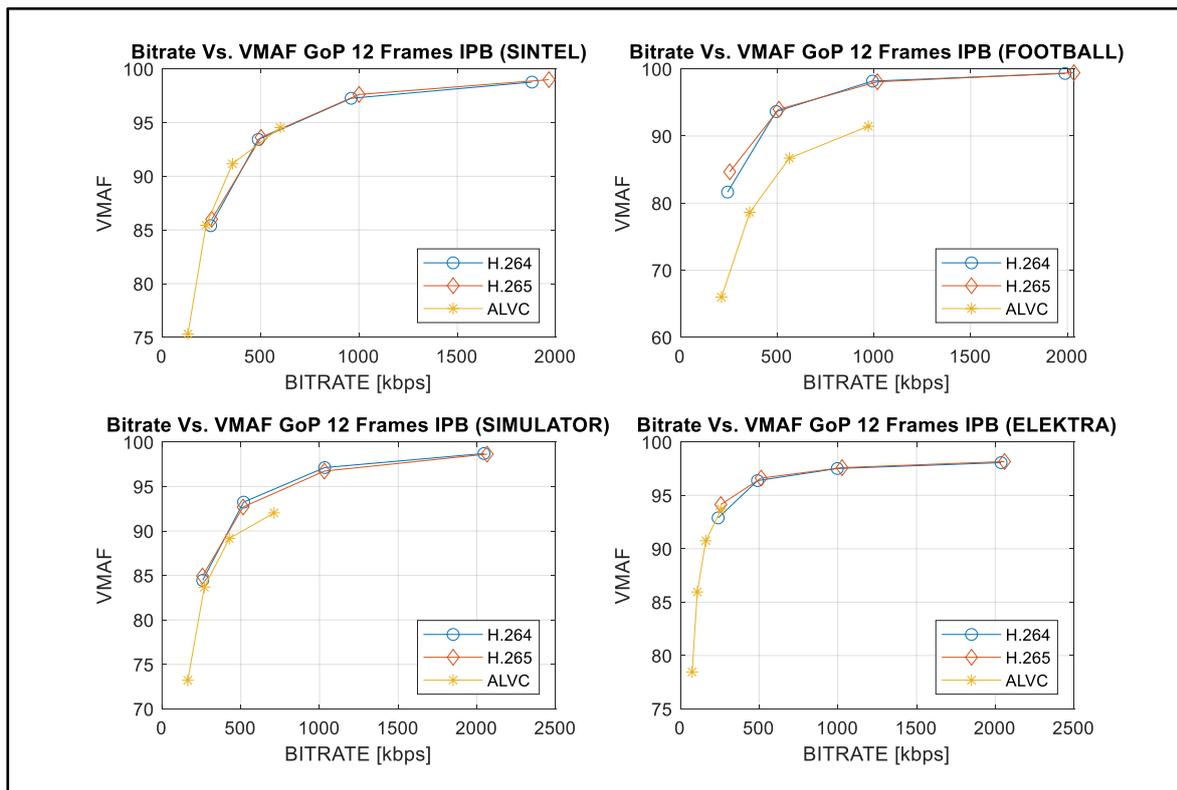


Fig. 24: Gráficas VMAF GoP 12 frames IPB.

III.2.6. Valores VMAF GoP 24 codificación frames tipo IPB

VMAF - GROUP OF PICTURES GOP 24 (IPB)							
VÍDEO	BITRATE [kbps]	CODIFICACIÓN TRADICIONAL				CODIFICACIÓN IA	
		Bitrate Real [kbps]	VMAF H.264	Bitrate Real [kbps]	VMAF H.265	Bitrate Real [kbps]	VMAF ALVC
SINTEL	2048	1846	98.8095	1905	98.9992	479.877	93.9899
	1024	927	97.5635	963	97.8409	289.583	90.4992
	512	463	94.3787	482	94.4767	201.764	84.4517
	256	229	87.0223	243	87.6880	122.843	75.6743
FOOTBALL	2048	1925	99.3160	2006	99.4722	817.325	91.1525
	1024	956	98.4300	997	98.4143	485.877	86.3302
	512	471	94.4301	498	94.8130	320.258	78.2420
	256	230	82.8175	238	84.8305	196.978	66.1191
SIMULATOR	2048	1930	98.7168	2026	98.6926	592.590	91.4969
	1024	979	97.2043	1004	96.8735	354.825	88.4687
	512	478	93.4902	507	93.3875	234.735	82.8067
	256	237	85.6511	256	86.6496	144.873	74.0125
ELEKTRA	2048	1939	98.0455	2047	98.2130	215.145	93.2233
	1024	936	97.5628	1007	97.6917	140.330	90.0692
	512	453	96.6093	492	96.7597	101.660	84.6430
	256	221	93.7836	241	94.6192	69.256	78.4079

Tabla 10: Valores VMAF GoP 24 codificación frames tipo IPB.

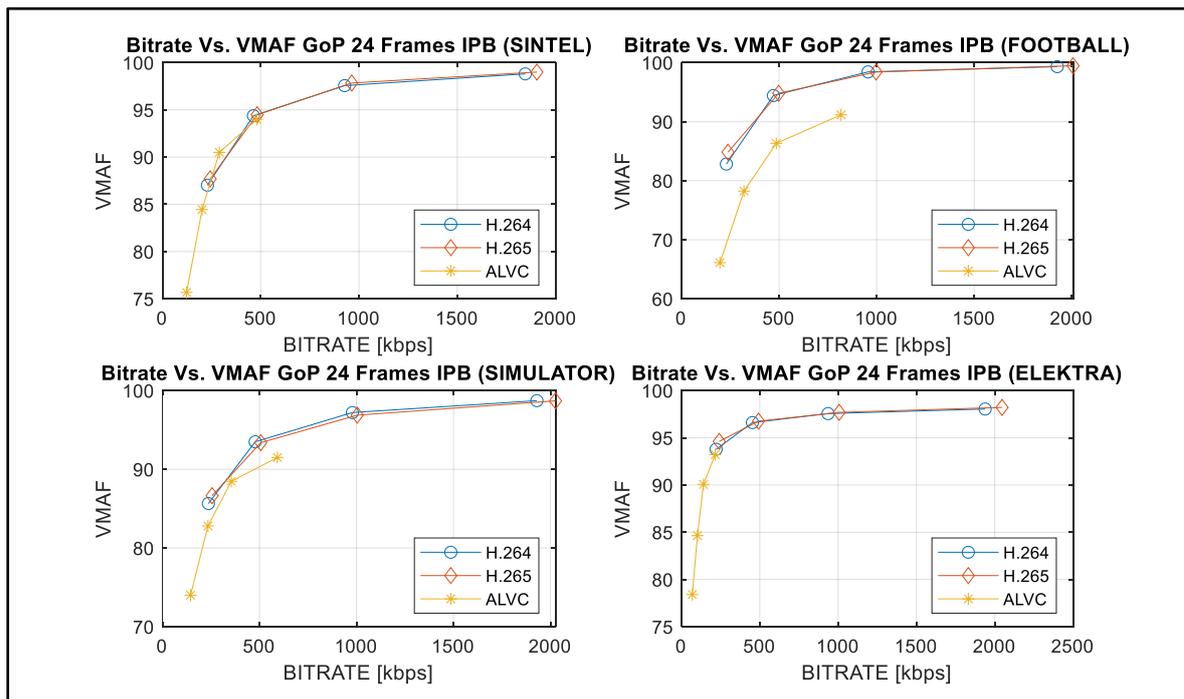


Fig. 25: Gráficas VMAF GoP 24 frames IPB.

IV. DISCUSIÓN

IV.1. DATOS RELEVANTES DEL ESTUDIO BIBLIOGRÁFICO

El estudio bibliográfico en su análisis cualitativo y cuantitativo permite identificar algunas de las publicaciones relevantes y que son consideradas como referente en el campo de la codificación de vídeo con inteligencia artificial y en las investigaciones actuales que abordan esta área de investigación. Examinando criterios como autor, palabras clave, términos empleados en el título y resumen, año de publicación, región geográfica, número de visualizaciones y citas, evolución, enfoques tradicionales y con IA para la compresión de vídeo podemos resaltar algunas publicaciones que destacan por su contenido y por brindar un panorama amplio en el marco de la codificación de vídeo con técnicas de inteligencia artificial.

La publicación “*Designs and Implementations in Neural Network-based Video*” publicada en el repositorio ARXIV destaca el uso de redes neuronales profundas (DNN) en dos niveles: el primero relacionado con la integración de herramientas de codificación basada en redes neuronales siguiendo la estructura de los estándares tradicionales; y el segundo que plantea la construcción de toda la estructura de compresión sobre redes neuronales. El artículo muestra la integración del primer enfoque empleando redes neuronales para la predicción *Intra* y el filtrado en bucle [18]. La predicción *Intra* implementada sobre redes neuronales (NN) realiza un mapeo sobre muestras vecinas para la predicción de un bloque actual con NN conectadas, y por otro lado, el filtrado en bucle se implementa con redes neuronales convolucionales (CNN) con el objetivo de recuperar los detalles perdidos durante la compresión [18].

El repositorio ACM presenta la publicación “*Deep Learning-Based Video Coding*” que muestra una breve revisión de los trabajos realizados en el campo de la codificación de vídeo utilizando técnicas de aprendizaje profundo (DL), al igual que la publicación destacada en el repositorio ARXIV, la publicación divide el estudio en dos categorías: la primera muestra el enfoque de nuevos esquemas de codificación sobre redes neuronales profundas (DNN) que muestra como aplicación el modelado de probabilidad de píxeles para la codificación predictiva y el codificador automático para la codificación de transformación; y la segunda categoría muestra las herramientas de codificación basadas en DNN aplicables a un marco de codificación tradicional, dentro de este enfoque se abordan técnicas de predicción de *Intraframe*, *Interframe*, canales, transformación, filtrado en bucle, muestreo ascendente y descendente [14]. La publicación además presenta un caso práctico denominado “*Deep Learning Video Coding (DLVC)*” que emplea dos herramientas basadas en CNN: el filtrado en bucle y la codificación de resolución adaptativa con un estudio de sus prestaciones en termino de PSNR comparados con estándares como H.264 y AV1 sobre una muestra de vídeos [14].

El artículo *“Image and Video Compression With Neural Networks: A Review”* del repositorio IEEE Xplore es un referente en el campo de la codificación de vídeo con IA que explora las prestaciones de las redes neuronales convolucionales (CNN), formadas por varias capas conectadas que pueden ser entrenadas masivamente por muestras de imágenes y vídeos, son vistas como extractores de características para transformar imágenes y vídeos, realizar tareas de clasificación, reconocimiento y predicción de datos que son características aplicables para la compresión de vídeo. La publicación introduce conceptos generales de compresión y redes neuronales, además de explorar otras técnicas para la codificación de vídeo utilizando CNN [71].

Science Direct con el artículo *“Machine learning based video coding optimizations: A survey”* se centra en como las técnicas del aprendizaje profundo (DL) con los algoritmos de aprendizaje automático son utilizados para abordar enfoques de optimización de la codificación de vídeo por sus características de aprendizaje a partir de datos masivos, descubrimiento de patrones y toma de decisiones que proporcionan predicciones que son empleadas en el estudio de este campo [90]. La publicación aborda en primer lugar las prestaciones y redundancias en los datos de vídeo para examinar cómo pueden ser aprovechados por los nuevos enfoques de codificación. Luego explora los requisitos y desafíos clave en la optimización de codificación de vídeo como son la relación de compresión, complejidad de codificación y calidad visual. Y por último, presenta la optimización de codificación de vídeo de alta eficiencia basada en aprendizaje que explica el enfoque de aplicación de las técnicas de aprendizaje profundo para abordar la codificación predictiva, la codificación de transformación y los algoritmos de mejora para incrementar la calidad del vídeo reconstruido [90].

El repositorio Springer aporta a este campo con la publicación *“Video super-resolution based on deep learning: a comprehensive survey”* que presenta un estudio de técnicas basada en redes neuronales profundas (DNN) para abordar un campo relacionado con la codificación de vídeo como es la superresolución de vídeo que tiene como objetivo la reconstrucción de vídeos de alta resolución a partir de otros de baja resolución. Si bien existen algoritmos tradicionales para abordar este enfoque, los métodos basados en aprendizaje profundo (DL) toman fuerza y se proponen soluciones basadas en redes GAN, CNN y RNN que tiene como principio fundamental emplear una gran cantidad de datos en términos de secuencias de vídeo en baja y alta resolución que ingresan a la red neuronal (NN) que está presente en cada uno de los módulos de alineación, extracción, fusión de características y reconstrucción que forman parte la arquitectura fundamental de las técnicas de superresolución de vídeo para aprovechar la información entre cuadros [106]. La superresolución de vídeo es un área en constante investigación por sus múltiples aplicaciones en la medicina, videovigilancia y televisión de alta definición.

IV.2. REVISIÓN DE LA EVALUACIÓN DE PRESTACIONES

La evaluación de las prestaciones de las soluciones propuestas de codificación artificial comienza con un análisis pre-codificación para cada uno de los vídeos que forman parte del banco de vídeos de prueba: Sintel, Football, Simulator y Elektra, para cada uno de los cuales se extrajo por medio de *ffmpeg* la información general de cada uno de ellos. En la Tabla 3 podemos observar una duración promedio de 10 segundos, un estándar de codificación H.264, un formato yuv420p, una resolución espacial de 416x240 y una resolución temporal de 24 fps que son parámetros comunes para todos los vídeos de prueba, el bitrate original de los vídeos es 8067, 10201, 7326, 7449 kbps para Sintel, Football, Simulator y Elektra respectivamente.

Analizar la información espacial (SI) y temporal (TI) nos permite obtener características sobre la complejidad del contenido de los vídeos. En la Tabla 4 se muestra la información extraída para cada uno de los vídeos de prueba utilizando *ffmpeg*, la información espacial representa la complejidad de los detalles visuales dentro de los fotogramas mientras que la información temporal muestra la cantidad de movimiento y cambios entre los fotogramas. En la Figura 19 se muestra la distribución y la densidad de los datos SITI para cada uno de los vídeos: Elektra presenta una baja complejidad espacial y temporal que sugiere un vídeo con poco detalle visual y movimiento por lo que la compresión puede ser muy eficiente; Sintel es un vídeo promedio en complejidad espacial y temporal; Football es el vídeo con mayor complejidad temporal que indica mayor movimiento y cambios rápidos entre fotogramas; y finalmente Simulator tiene el valor más alto de complejidad espacial por lo que presenta muchos detalles visuales en sus fotogramas.

La interfaz de ejecución para probar los modelos de los codificadores IA (ENVC y ALVC), se implementa en un Sistema Operativo Windows sobre un entorno de Linux, utilizando *Windows Subsystem for Linux* (WSL) para ejecutar un entorno CONDA, donde se instalaron los paquetes/librerías de Python y aprendizaje automático como PyTorch y TensorFlow. Los entornos creados utilizan los recursos de la tarjeta gráfica NVIDIA GeForce MX150, permitiendo ejecutar y procesar los modelos de entrenamiento y operaciones de aprendizaje profundo de las soluciones IA. Los recursos de la tarjeta gráfica pueden verse limitados para procesar videos de resoluciones mayores, además de requerir un alto tiempo de ejecución para procesar los fotogramas de los videos de prueba, por lo que se emplearon varias semanas para realizar las pruebas y obtener resultados.

Para evaluar las prestaciones de los codificadores tradicionales y los basados en IA se analizaron los resultados obtenidos para la métrica objetiva PSNR y la métrica perceptual VMAF codificando los vídeos de prueba con H.264/H.265 (Codificación tradicional) y ENVC/ALVC (Codificación IA). Para un primer análisis se ajustaron los parámetros de codificación a un GoP tamaño 12 con tramas IP, se realizaron codificaciones con H.264, H.265, ENVC y ALVC con bitrates entre 256 y 4096 kbps obteniendo los valores de Bitrate Real [Kbps] y PSNR [dB] que se muestran en la Tabla 5 y Figura 20. Se puede observar que generalmente H.265 ofrece una mejor calidad con valores de PSNR

más altos a un bitrate ligeramente menor que H.264 que se debe a una mayor eficiencia de compresión de H.265 que busca reducir el tamaño del archivo manteniendo o mejorando la calidad del vídeo. Por otra parte, la codificación basada en IA (ENVC/ALVC) tiene valores de bitrate real significativamente menores que los estándares tradicionales lo que muestra una eficiencia mayor en la compresión de los vídeos, pero con una penalización en la calidad al presentar valores de PSNR considerablemente más bajos que H.264 y H.265. Al comparar los resultados entre la codificación IA podemos observar que ENVC conserva un mejor balance en términos de compresión y calidad con valores de bitrates más bajos y valores ligeramente mayores de PSNR frente a ALVC.

Como segundo análisis se ajustaron los parámetros de codificación a un GoP tamaño 24 con tramas IP, se realizaron codificaciones con H.264, H.265, ENVC y ALVC con bitrates entre 256 y 4096 kbps obteniendo los valores de Bitrate Real [Kbps] y PSNR [dB] que se muestran en la Tabla 6 y Figura 21. Se observan resultados similares al primer análisis que muestran superioridad de H.265 en calidad con valores de PSNR más altos y en compresión con un bitrate real menor que H.264, la codificación IA (ENVC/ALVC) sigue manteniendo valores de bitrate real mucho menores mostrando mayor eficiencia en términos de compresión, sin embargo, en términos de calidad presenta medidas de PSNR bajos. Al comparar ENVC y ALVC con esta nueva configuración observamos un comportamiento similar a la configuración del primer análisis con una superioridad de ENVC frente ALVC en términos de compresión y calidad. La configuración de un GoP tamaño 24 en general permite una codificación más eficiente en términos de compresión, la razón se debe a que utilizando GoP largos se codifican menos frames I en el vídeo por lo que el bitrate real es menor mejorando la compresión y permitiendo utilizar menor ancho de banda de la red para transmitir mayor contenido multimedia.

En un tercer análisis se ajustaron los parámetros de codificación a un GoP tamaño 12 con tramas IPB, y se realizaron codificaciones con H.264, H.265 y ALVC con bitrates entre 256 y 2048 kbps obteniendo los valores de Bitrate Real [Kbps] y PSNR [dB] que se muestran en la Tabla 7 y Figura 22 (para este estudio no se analiza las prestaciones de ENVC puesto que su modelo no permite la incorporación de tramas tipo B). Adicionalmente se utiliza un rango de bitrate entre 256 y 2048 kbps ya que los modelos de entrenamiento de ALVC permiten codificar únicamente en este rango de valores. La incorporación de tramas B en el proceso de codificación representa más retos sobre todo para la codificación IA porque es computacionalmente más complejo, tiene un mayor retardo puesto que requiere recibir las tramas P o I para la codificación o decodificación de las tramas B, necesita mayores recursos como mayor memoria de procesamiento y almacenamiento, pero sin embargo, ofrece ventajas en términos de eficiencia de compresión al incrementarse el ratio de compresión de los vídeos, ya que las tramas B ocupan menor espacio que las tramas I y P. Esto se ve reflejado en los valores del bitrate real al presentar una disminución para H.264, H.265 y ALVC en comparación con los resultados del primer análisis. Así mismo, se observa una mejora en términos de calidad al

incrementarse los valores de PSNR por el aporte que las tramas B brindan en el proceso de codificación mejorando la predicción para reconstruir el vídeo con una mayor precisión.

El cuarto análisis consistió en ajustar los parámetros de codificación a un GoP tamaño 24 con tramas IPB, se realizaron codificaciones con H.264, H.265 y ALVC con bitrates entre 256 y 2048 kbps obteniendo los valores de Bitrate Real [Kbps] y PSNR [dB] que se muestran en la Tabla 8 y Figura 23. Los resultados muestran una reducción en el bitrate real de H.264, H.265 y ALVC que se expresa en el aumento de la eficiencia de compresión manteniendo una calidad estable que se ve reflejada en algunos casos en el aumento del PSNR y en otros manteniéndolo con respecto a los análisis realizados anteriormente.

El quinto y sexto análisis tienen como objetivo obtener la medida de la métrica de calidad visual perceptual VMAF desarrollada por Netflix y expresada en la escala entre 0 y 100, se obtuvieron los valores utilizando la librería VMAF que incorpora FFMPEG para la extracción de estos parámetros para los vídeos codificados con H.264, H.265 y ALVC con las configuraciones de GoP tamaño 12 y 24 con tramas IPB para los bitrates de 256, 512, 1024 y 2048 kbps. Las Tablas 9 y 10 junto con las Figuras 24 y 25 muestran estos resultados. Como análisis general en su gran mayoría se observan mejores resultados para la configuración de GoP tamaño 24, los valores más bajos de la medida VMAF se encuentran en el bitrate de 256 kbps en ambas configuraciones, 65 para la codificación IA y 82 para la codificación tradicional. Sin embargo, a medida que el bitrate aumenta la medida crece llegando a alcanzar incluso valores de 99 para la codificación tradicional y valores hasta de 94 para la codificación IA para el bitrate más alto. Si bien VMAF nos brinda una medida de la calidad visual de los vídeos y la codificación tradicional tiene mejores resultados en las dos configuraciones frente a la codificación IA, se puede establecer que ambas soluciones aportan con una buena calidad de vídeo para los bitrates bajos que implican ligeras degradaciones perceptibles pero aceptables para los usuarios que no afectan a la experiencia visual y una excelente calidad de vídeo para los bitrates más altos que sugiere alto grado de similitud al vídeo de referencia por lo que las degradaciones que puedan existir se consideran insignificantes y prácticamente son indistinguibles por el usuario.

IV.3. LINEAS FUTURAS DE INVESTIGACIÓN

Reducción de la complejidad computacional para una compresión eficiente: la codificación tradicional se ha desarrollado continuamente y sus algoritmos han permitido que factores como el tiempo de codificación alcance valores aceptables. Sin embargo, en la codificación IA la complejidad computacional que se requiere para procesar las redes profundas convierte al tiempo de codificación en un parámetro crítico, el estudio y desarrollo de redes neuronales más compactas y eficientes permitirán que la codificación IA sea posible en escenarios con limitados recursos computacionales para lograr una compresión eficiente.

Esquemas completos de codificación sobre redes neuronales profundas (DNN): si bien el estudio del campo de la codificación de vídeo con IA ha empezado con el desarrollo e incorporación de herramientas profundas que se combinan con los enfoques tradicionales de codificación creando esquemas híbridos, el siguiente paso es desarrollar esquemas completos de codificación sobre DNN que alcancen y superen a los enfoques tradicionales brindando mayores posibilidades en términos de eficiencia de compresión y calidad visual.

Algoritmos de codificación IA para superresolución de vídeo: la superresolución de vídeo permitirá el desarrollo de aplicaciones de videovigilancia, realidad virtual y aumentada, visualización de alta resolución superiores a 8K, por lo que un desafío es el desarrollo de algoritmos profundos de superresolución de vídeo más eficientes, con menos parámetros y que requieran menos recursos computacionales y de almacenamiento para su entrenamiento.

V. CONCLUSIONES

El estudio bibliográfico sobre la codificación de vídeo con IA mostró que China y Estados Unidos son los países que más aportan científicamente, 2023 es el año con más publicaciones en todos los repositorios, IEEE Xplore cuenta con más artículos sobre este campo, las palabras clave que se utilizan con más frecuencia son: *video compression*, *video coding* y *learned video compression*, así mismo los términos más empleados en títulos y resúmenes son: *neuronal representation*, *machine learning*, *deep neuronal network* y *video super resolution*. La codificación de vídeo ha experimentado una constante evolución desde 1984 con la aparición del primer estándar de códec de vídeo (H.120) hasta llegar al último estándar de códec de vídeo presentado en 2021 (H.266), en la actualidad la investigación se enfoca en el estudio de herramientas y sistemas basados en aprendizaje profundo (DL) que tienen como unidad central a las redes neuronales profundas (DNN), desde el 2020 los artículos han empezado a incorporar en sus publicaciones acceso a código abierto de algoritmos, modelos de entrenamiento y códecs para realizar pruebas e investigación.

La evaluación de prestaciones de las soluciones propuestas mostró que los estándares tradicionales como H.264 y H.265 presentan mayor eficiencia en términos de calidad visual con valores más altos en las medidas de PSNR frente a la codificación IA, por otro lado, la codificación IA con ENVC y ALVC alcanzan una mayor eficiencia en términos de compresión al reducir significativamente el resultado del bitrate real que se traduce en una reducción considerable en el tamaño del archivo, aquellos vídeos con mayor complejidad espacial o temporal pueden requerir más recursos para una codificación eficiente frente aquellos que cuentan con poco detalle visual y movimiento para los cuales la compresión puede ser muy eficiente.

Las medidas obtenidas para la métrica de calidad visual perceptual VMAF realizada a los vídeos codificados con H.264/H.265(Codificación tradicional) y ALVC (Codificación IA), muestran que si bien la codificación IA tiene valores más bajos que la codificación tradicional, ambas soluciones aportan con una buena calidad de vídeo, en los bitrates bajos existen ligeras degradaciones

perceptibles pero aceptables para los usuarios que no afectan a la experiencia visual y en los bitrates altos se observa una excelente calidad de vídeo que sugiere alto grado de similitud al vídeo de referencia con degradaciones insignificantes e indistinguibles por el usuario.

En el proceso de codificación la variación del tamaño de GoP y la incorporación de tramas tipo B a la estructura puede aportar con mejores resultados en términos de compresión y calidad visual, utilizar periodos de GoP largos reduce el bitrate real del vídeo codificado que puede ser útil en la transmisión de contenido multimedia logrando una optimización del ancho de banda de la red, para aplicaciones que requieran almacenamiento de contenido multimedia es útil puesto que se reduce el tamaño final del archivo. Si bien es cierto que la incorporación de tramas B en el proceso de codificación es computacionalmente más complejo y se requiere más memoria de almacenamiento también se incrementa el ratio de compresión ya que ocupan menos espacio que tramas tipo I y P.

AGRADECIMIENTOS

Este trabajo va dedicado a Dios por ser mi guía, bendición, refugio y fortaleza, a mis padres por su amor y apoyo incondicional presentes en todas las etapas de mi vida. Un agradecimiento sincero a la Universidad Politécnica de Valencia, la asociación AUIP, al MUTSRC y a Juan Carlos Guerri por dirigir este trabajo junto a Pau Arce Vila. El TFM se ha realizado con la colaboración del grupo de investigación COMM (*Comunicaciones Multimedia*) del iTEAM en el marco del proyecto CDTI IDI-20240021 “*Desarrollo de dispositivos de visualización inteligente para medios de transporte*”.

BIBLIOGRAFÍA BASE

- [1] W. Gao, S. Sun, H. Zheng, Y. Wu, H. Ye, and Y. Zhang, “OpenDMC: An Open-Source Library and Performance Evaluation for Deep-learning-based Multi-frame Compression,” *MM 2023 - Proceedings of the 31st ACM International Conference on Multimedia*, pp. 9685–9688, Oct. 2023, doi: 10.1145/3581783.3613464.
- [2] H. Gao, J. Cui, M. Ye, S. Li, Y. Zhao, and X. Zhu, “Structure-Preserving Motion Estimation for Learned Video Compression,” *MM 2022 - Proceedings of the 30th ACM International Conference on Multimedia*, pp. 3055–3063, Oct. 2022, doi: 10.1145/3503161.3548156.
- [3] J. Han, S. Lombardo, C. S. Disneyresearch|studios, and S. Mandt, “Deep Generative Video Compression,” *NIPS’19 - 33rd International Conference on Neural Information Processing Systems*, 2019, doi: 10.5555/3454287.3455120.
- [4] J. Li, B. Li, and Y. Lu, “Hybrid Spatial-Temporal Entropy Modelling for Neural Video Compression,” *MM 2022 - Proceedings of the 30th ACM International Conference on Multimedia*, pp. 1503–1511, Oct. 2022, doi: 10.1145/3503161.3547845.
- [5] Y. Shi, S. An, J. Wang, and B. Yin, “Learned Bi-Directional Motion Prediction for Video Compression,” *Proceedings of the 4th ACM International Conference on Multimedia in Asia, MMAsia 2022*, Dec. 2022, doi: 10.1145/3551626.3564954.
- [6] M. Li, Y. Shi, J. Wang, and Y. Huang, “High Visual-Fidelity Learned Video Compression,” *MM 2023 - Proceedings of the 31st ACM International Conference on Multimedia*, pp. 8057–8066, Oct. 2023, doi: 10.1145/3581783.3612530.
- [7] Y. Hu *et al.*, “Deep Inter Prediction with Error-Corrected Auto-Regressive Network for Video Coding,” *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 19, no. 1, Jan. 2023, doi: 10.1145/3528173.
- [8] H. Lin, B. Chen, Z. Zhang, J. Lin, X. Wang, and T. Zhao, “DeepSVC: Deep Scalable Video Coding for Both Machine and Human Vision,” *MM 2023 - Proceedings of the 31st ACM International Conference on Multimedia*, pp. 9205–9214, Oct. 2023, doi: 10.1145/3581783.3612500.

- [9] H. Yuan and W. Gao, "OpenFastVC: An Open Source Library for Video Coding Fast Algorithm Implementation," *MM 2023 - Proceedings of the 31st ACM International Conference on Multimedia*, pp. 9660–9663, Oct. 2023, doi: 10.1145/3581783.3613465.
- [10] T. Zhao, W. Feng, H. J. Zeng, Y. Xu, Y. Niu, and J. Liu, "Learning-Based Video Coding with Joint Deep Compression and Enhancement," *MM 2022 - Proceedings of the 30th ACM International Conference on Multimedia*, pp. 3045–3054, Oct. 2022, doi: 10.1145/3503161.3548314.
- [11] HuoShuai *et al.*, "Towards Hybrid-Optimization Video Coding," *ACM Comput Surv*, Aug. 2022, doi: 10.1145/3652148.
- [12] M. Shukor, B. B. Damodaran, X. Yao, and P. Hellier, "Video Coding using Learned Latent GAN Compression," *MM 2022 - Proceedings of the 30th ACM International Conference on Multimedia*, pp. 2239–2248, Oct. 2022, doi: 10.1145/3503161.3548219.
- [13] ZhangTicao and MaoShiwen, "An Overview of Emerging Video Coding Standards," *GetMobile: Mobile Computing and Communications*, vol. 22, no. 4, pp. 13–20, May 2019, doi: 10.1145/3325867.3325873.
- [14] L. I. U. Dong, L. I. Yue, L. I. N. Jianping, L. I. Houqiang, and W. U. Feng, "Deep Learning-Based Video Coding," *ACM Computing Surveys (CSUR)*, vol. 53, no. 1, Feb. 2020, doi: 10.1145/3368405.
- [15] H. Chen, M. Gwilliam, S.-N. Lim, and A. Shrivastava, "HNerV: A Hybrid Neural Representation for Videos," pp. 10270–10279, Apr. 2023, doi: 10.1109/cvpr52729.2023.00990.
- [16] H. M. Kwan, G. Gao, F. Zhang, A. Gower, and D. Bull, "HiNerV: Video Compression with Hierarchical Encoding-based Neural Representation," Jun. 2023, Accessed: Apr. 11, 2024. [Online]. Available: <https://arxiv.org/abs/2306.09818v3>
- [17] H. Hadizadeh and I. V. Bajić, "Learned Scalable Video Coding For Humans and Machines," Jul. 2023, Accessed: Apr. 11, 2024. [Online]. Available: <http://arxiv.org/abs/2307.08978>
- [18] Y. Li *et al.*, "Designs and Implementations in Neural Network-based Video Coding," Sep. 2023, Accessed: Apr. 11, 2024. [Online]. Available: <https://arxiv.org/abs/2309.05846v2>
- [19] C. Jia *et al.*, "MPAI-EEV: Standardization Efforts of Artificial Intelligence based End-to-End Video Coding," Sep. 2023, Accessed: Apr. 11, 2024. [Online]. Available: <http://arxiv.org/abs/2309.07589>
- [20] S. Zhang, W. Mao, H. Shi, and Z. Wang, "A Computationally Efficient Neural Video Compression Accelerator Based on a Sparse CNN-Transformer Hybrid Network," Dec. 2023, Accessed: Apr. 11, 2024. [Online]. Available: <http://arxiv.org/abs/2312.10716>
- [21] Z. Li *et al.*, "A Video Coding Method Based on Neural Network for CLIC2024," Jan. 2024, Accessed: Apr. 11, 2024. [Online]. Available: <https://arxiv.org/abs/2401.03623v1>
- [22] T. Leguay, T. Ladune, P. Philippe, and O. Déforges, "Cool-chic video: Learned video coding with 800 parameters," Feb. 2024, Accessed: Apr. 11, 2024. [Online]. Available: <https://arxiv.org/abs/2402.03179v2>
- [23] O. G. Guleryuz *et al.*, "Sandwiched Compression: Repurposing Standard Codecs with Neural Network Wrappers," Feb. 2024, Accessed: Apr. 11, 2024. [Online]. Available: <https://arxiv.org/abs/2402.05887v1>
- [24] Y. Zhao *et al.*, "A Neural-network Enhanced Video Coding Framework beyond ECM," Feb. 2024, Accessed: Apr. 11, 2024. [Online]. Available: <http://arxiv.org/abs/2402.08397>
- [25] T. van Rozendaal, I. A. M. Huijben, and T. S. Cohen, "Overfitting for Fun and Profit: Instance-Adaptive Data Compression," *ICLR 2021 - 9th International Conference on Learning Representations*, Jan. 2021, Accessed: Apr. 11, 2024. [Online]. Available: <http://arxiv.org/abs/2101.08687>
- [26] J. Bégaint, F. Racapé, S. Feltman, and A. Pushparaja, "CompressAI: a PyTorch library and evaluation platform for end-to-end compression research," Nov. 2020, Accessed: Apr. 10, 2024. [Online]. Available: <http://arxiv.org/abs/2011.03029>
- [27] A. K. R. D. and N. Udupa, "A Study of the Evolution of Video Codec and its Future Research Direction," in *2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAIECC)*, IEEE, Dec. 2020, pp. 1–13. doi: 10.1109/ICAIECC50550.2020.9339513.
- [28] J. Li, B. Li, and Y. Lu, "Neural Video Compression with Diverse Contexts," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2023, pp. 22616–22626. doi: 10.1109/CVPR52729.2023.02166.
- [29] M. Khani, V. Sivaraman, and M. Alizadeh, "Efficient Video Compression via Content-Adaptive Super-Resolution," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2021, pp. 4501–4510. doi: 10.1109/ICCV48922.2021.00448.
- [30] G. He *et al.*, "A Video Compression Framework Using an Overfitted Restoration Neural Network," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, Jun. 2020, pp. 593–597. doi: 10.1109/CVPRW50498.2020.00082.
- [31] G. He *et al.*, "An Efficient Video Coding System With an Adaptive Overfitted Multi-Scale Attention Network," *IEEE Access*, vol. 9, pp. 64022–64032, 2021, doi: 10.1109/ACCESS.2021.3075623.

- [32] D. Ding, Z. Ma, D. Chen, Q. Chen, Z. Liu, and F. Zhu, “Advances in Video Compression System Using Deep Neural Network: A Review and Case Studies,” *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1494–1520, Sep. 2021, doi: 10.1109/JPROC.2021.3059994.
- [33] D. P. S. Subramanian, V. Suresh, R. Shivalli, and S. Sinha, “Video Compression using Deep Neural Networks,” in *2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP)*, IEEE, Dec. 2022, pp. 1–5. doi: 10.1109/CCIP57447.2022.10058645.
- [34] M. C. de Oliveira *et al.*, “Learning-based End-to-End Video Compression Using Predictive Coding,” in *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, IEEE, Oct. 2021, pp. 160–167. doi: 10.1109/SIBGRAPI54419.2021.00030.
- [35] A. Punchihewa and D. Bailey, “A Review of Emerging Video Codecs: Challenges and Opportunities,” in *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, IEEE, Nov. 2020, pp. 1–6. doi: 10.1109/IVCNZ51579.2020.9290536.
- [36] J. Pessoa, H. Aidos, P. Tomas, and M. A. T. Figueiredo, “End-to-End Learning of Video Compression using Spatio-Temporal Autoencoders,” in *2020 IEEE Workshop on Signal Processing Systems (SiPS)*, IEEE, Oct. 2020, pp. 1–6. doi: 10.1109/SiPS50750.2020.9195249.
- [37] X. Zhang, J. Shao, and J. Zhang, “Low-complexity Deep Video Compression with A Distributed Coding Architecture,” in *2023 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, Jul. 2023, pp. 2537–2542. doi: 10.1109/ICME55011.2023.00432.
- [38] H. Lin, X. He, L. Qing, Q. Teng, and S. Yang, “Improved Low-Bitrate HEVC Video Coding Using Deep Learning Based Super-Resolution and Adaptive Block Patching,” *IEEE Trans Multimedia*, vol. 21, no. 12, pp. 3010–3023, Dec. 2019, doi: 10.1109/TMM.2019.2919433.
- [39] H. Man, C. Yu, F. Xing, Y. Cheng, B. Zheng, and X. Fan, “Deep Learning-Assisted Video Compression Framework,” in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, May 2022, pp. 3210–3214. doi: 10.1109/ISCAS48785.2022.9937452.
- [40] W.-H. Peng and H.-M. Hang, “Recent Advances in End-to-End Learned Image and Video Compression,” in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, IEEE, Dec. 2020, pp. 1–2. doi: 10.1109/VCIP49819.2020.9301753.
- [41] B. Liu, Y. Chen, S. Liu, and H.-S. Kim, “Deep Learning in Latent Space for Video Prediction and Compression,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2021, pp. 701–710. doi: 10.1109/CVPR46437.2021.00076.
- [42] F. Zhang, M. Afonso, and D. R. Bull, “Enhanced Video Compression Based on Effective Bit Depth Adaptation,” in *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE, Sep. 2019, pp. 1720–1724. doi: 10.1109/ICIP.2019.8803185.
- [43] Y. Wei, L. Chen, and L. Song, “Video Compression based on Jointly Learned Down-Sampling and Super-Resolution Networks,” in *2021 International Conference on Visual Communications and Image Processing (VCIP)*, IEEE, Dec. 2021, pp. 1–5. doi: 10.1109/VCIP53242.2021.9675356.
- [44] J. Yang, C. Yang, F. Xiong, F. Wang, and R. Wang, “Learned Low Bitrate Video Compression with Space-Time Super-Resolution,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, Jun. 2022, pp. 1785–1789. doi: 10.1109/CVPRW56347.2022.00192.
- [45] S. V. Pranav, P. Suraj, N. Pothineni, K. V. Kadambari, and P. Balasubramanyam, “Optimizing Video Compression & Decompression Using Deep Learning Networks,” in *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, IEEE, Sep. 2021, pp. 1–6. doi: 10.1109/GUCON50781.2021.9573656.
- [46] A. Jadhav, “Variable Rate Video Compression using a Hybrid Recurrent Convolutional Learning Framework,” in *2020 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, Jan. 2020, pp. 1–6. doi: 10.1109/ICCCI48352.2020.9104085.
- [47] D. Liu, Z. Chen, S. Liu, and F. Wu, “Deep Learning-Based Technology in Responses to the Joint Call for Proposals on Video Compression With Capability Beyond HEVC,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 5, pp. 1267–1280, May 2020, doi: 10.1109/TCSVT.2019.2945057.
- [48] D. Alexandre, H.-M. Hang, W.-H. Peng, and M. Domanski, “Deep Video Compression for Interframe Coding,” in *2021 IEEE International Conference on Image Processing (ICIP)*, IEEE, Sep. 2021, pp. 2124–2128. doi: 10.1109/ICIP42928.2021.9506275.
- [49] Y. Hu, C. Jung, Q. Qin, J. Han, Y. Liu, and M. Li, “HDVC: Deep Video Compression With Hyperprior-Based Entropy Coding,” *IEEE Access*, vol. 12, pp. 17541–17551, 2024, doi: 10.1109/ACCESS.2024.3350643.
- [50] S. Zhang, M. Mrak, L. Herranz, M. G. Blanch, S. Wan, and F. Yang, “DVC-P: Deep Video Compression with Perceptual Optimizations,” in *2021 International Conference on Visual*

- Communications and Image Processing (VCIP)*, IEEE, Dec. 2021, pp. 1–5. doi: 10.1109/VCIP53242.2021.9675350.
- [51] M. Khadir and M. F. Hashmi, “High Efficient Quality of Video Compression using Variational Autoencoders in Deep Learning,” in *2023 5th International Conference on Power, Control & Embedded Systems (ICPCES)*, IEEE, Jan. 2023, pp. 1–6. doi: 10.1109/ICPCES57104.2023.10075897.
- [52] Z. Chen, T. He, X. Jin, and F. Wu, “Learning for Video Compression,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 2, pp. 566–576, Feb. 2020, doi: 10.1109/TCSVT.2019.2892608.
- [53] D. Xu, G. Lu, R. Yang, and R. Timofte, “Learned image and video compression with deep neural networks,” in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, IEEE, Dec. 2020, pp. 1–3. doi: 10.1109/VCIP49819.2020.9301828.
- [54] Z. Zhang, Y. Li, K. Zhang, L. Zhang, and Y. He, “Learning-Based End-to-End Video Compression with Spatial-Temporal Adaptation,” in *2022 IEEE International Conference on Image Processing (ICIP)*, IEEE, Oct. 2022, pp. 2821–2825. doi: 10.1109/ICIP46576.2022.9897830.
- [55] D. Jin, J. Lei, B. Peng, Z. Pan, L. Li, and N. Ling, “Learned Video Compression With Efficient Temporal Context Learning,” *IEEE Transactions on Image Processing*, vol. 32, pp. 3188–3198, 2023, doi: 10.1109/TIP.2023.3276333.
- [56] C. Liu, H. Sun, X. Zeng, and Y. Fan, “Learned Video Compression With Residual Prediction And Feature-Aided Loop Filter,” in *2022 IEEE International Conference on Image Processing (ICIP)*, IEEE, Oct. 2022, pp. 1321–1325. doi: 10.1109/ICIP46576.2022.9897989.
- [57] M. Chen, A. Patney, and A. C. Bovik, “MOVI-Codec: Deep Video Compression without Motion,” in *2021 Picture Coding Symposium (PCS)*, IEEE, Jun. 2021, pp. 1–5. doi: 10.1109/PCS50896.2021.9477463.
- [58] A. Maksimov and M. Gashnikov, “Generalization of Machine Learning-Based Image Compression Methods for Video Compression,” in *2023 IX International Conference on Information Technology and Nanotechnology (ITNT)*, IEEE, Apr. 2023, pp. 1–5. doi: 10.1109/ITNT57377.2023.10139119.
- [59] M. A. Yakubenko and M. V. Gashnikov, “Entropy Modeling in Video Compression Based on Machine Learning,” in *2023 IX International Conference on Information Technology and Nanotechnology (ITNT)*, IEEE, Apr. 2023, pp. 1–4. doi: 10.1109/ITNT57377.2023.10139143.
- [60] M. Khadir, M. F. Hashmi, and M. Khadir, “Deep Learning Empowers Next-Gen Video Compression: Bridging Quality and Efficiency,” in *2023 24th International Arab Conference on Information Technology (ACIT)*, IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ACIT58888.2023.10453895.
- [61] H. Wang and Z. Chen, “Exploring Long- and Short-Range Temporal Information for Learned Video Compression,” *IEEE Transactions on Image Processing*, vol. 33, pp. 780–792, 2024, doi: 10.1109/TIP.2024.3349859.
- [62] Y. Wang, Q. Huang, B. Tang, W. Liu, W. Shan, and Q. Xu, “Learned Video Compression with Spatial-Temporal Optimization,” *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3715–3719, Apr. 2024, doi: 10.1109/ICASSP48485.2024.10446198.
- [63] H. M. Cho and K. Choi, “Super-Resolution based Video Coding Scheme,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, Jun. 2022, pp. 1777–1779. doi: 10.1109/CVPRW56347.2022.00190.
- [64] J. Yang, C. Yang, Y. Zhai, Q. Wang, X. Pan, and R. Wang, “Improving Learned Video Compression by Exploring Spatial Redundancy,” in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Apr. 2024, pp. 2860–2864. doi: 10.1109/ICASSP48485.2024.10448496.
- [65] A. Ibaba, S. Adeshina, and A. M. Aibinu, “A Review of Video Compression Optimization Techniques,” in *2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS)*, IEEE, Jul. 2021, pp. 1–5. doi: 10.1109/ICMEAS52683.2021.9692416.
- [66] D. Petreski and T. Kartalov, “Next Generation Video Compression Standards – Performance Overview,” in *2023 30th International Conference on Systems, Signals and Image Processing (IWSSIP)*, IEEE, Jun. 2023, pp. 1–5. doi: 10.1109/IWSSIP58668.2023.10180261.
- [67] A. B. Bräscher, G. F. Da Silveira, L. H. Cancellier, I. Seidel, M. Grellert, and J. L. Güntzel, “A New Approach to Video Coding Leveraging Hybrid Coding and Video Frame Interpolation,” in *2023 36th SBC/SBMicro/IEEE/ACM Symposium on Integrated Circuits and Systems Design (SBCCI)*, IEEE, Aug. 2023, pp. 1–6. doi: 10.1109/SBCCI60457.2023.10261663.
- [68] L. Tang, X. Zhang, G. Zhang, and X. Ma, “Scene Matters: Model-based Deep Video Compression,” in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2023, pp. 12447–12457. doi: 10.1109/ICCV51070.2023.01147.

- [69] M. A. Chubar and M. V. Gashnikov, "Deep Contextual Video Compression Based on Machine Learning," in *2023 IX International Conference on Information Technology and Nanotechnology (ITNT)*, IEEE, Apr. 2023, pp. 1–4. doi: 10.1109/ITNT57377.2023.10139047.
- [70] G. Lu, X. Zhang, W. Ouyang, L. Chen, Z. Gao, and D. Xu, "An End-to-End Learning Framework for Video Compression," *IEEE Trans Pattern Anal Mach Intell*, vol. 43, no. 10, pp. 3292–3308, Oct. 2021, doi: 10.1109/TPAMI.2020.2988453.
- [71] S. Ma, X. Zhang, C. Jia, Z. Zhao, S. Wang, and S. Wang, "Image and Video Compression With Neural Networks: A Review," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 6, pp. 1683–1698, Jun. 2020, doi: 10.1109/TCSVT.2019.2910119.
- [72] Z. Montajabi, V. Khorasani Ghassab, and N. Bouguila, "Recurrent Neural Network-Based Video Compression," in *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, Dec. 2022, pp. 925–930. doi: 10.1109/ICMLA55696.2022.00154.
- [73] A. Jacob, V. Pawar, V. Vishwakarma, and A. Mane, "Deep Learning Approach to Video Compression," in *2019 IEEE Bombay Section Signature Conference (IBSSC)*, IEEE, Jul. 2019, pp. 1–5. doi: 10.1109/IBSSC47189.2019.8973035.
- [74] O. Rippel, S. Nair, C. Lew, S. Branson, A. Anderson, and L. Bourdev, "Learned Video Compression," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2019, pp. 3453–3462. doi: 10.1109/ICCV.2019.00355.
- [75] R. Yang, R. Timofte, and L. Van Gool, "Advancing Learned Video Compression With In-Loop Frame Prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 5, pp. 2410–2423, May 2023, doi: 10.1109/TCSVT.2022.3222418.
- [76] H. M. Yasin and S. Y. Ameen, "Review and Evaluation of End-to-End Video Compression with Deep-Learning," in *2021 International Conference of Modern Trends in Information and Communication Technology Industry (MTICTI)*, IEEE, Dec. 2021, pp. 1–8. doi: 10.1109/MTICTI53925.2021.9664790.
- [77] Z. Guo, R. Feng, Z. Zhang, X. Jin, and Z. Chen, "Learning Cross-Scale Weighted Prediction for Efficient Neural Video Compression," *IEEE Transactions on Image Processing*, vol. 32, pp. 3567–3579, 2023, doi: 10.1109/TIP.2023.3287495.
- [78] T. Ladune and P. Philippe, "AIVC: Artificial Intelligence Based Video Codec," in *2022 IEEE International Conference on Image Processing (ICIP)*, IEEE, Oct. 2022, pp. 316–320. doi: 10.1109/ICIP46576.2022.9897240.
- [79] R. Yang, F. Mentzer, L. Van Gool, and R. Timofte, "Learning for Video Compression With Recurrent Auto-Encoder and Recurrent Probability Model," *IEEE J Sel Top Signal Process*, vol. 15, no. 2, pp. 388–401, Feb. 2021, doi: 10.1109/JSTSP.2020.3043590.
- [80] S. Zhu, C. Liu, and Z. Xu, "High-Definition Video Compression System Based on Perception Guidance of Salient Information of a Convolutional Neural Network and HEVC Compression Domain," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 1–1, Jul. 2020, doi: 10.1109/TCSVT.2019.2911396.
- [81] Z. Zhang, Y. Li, K. Zhang, L. Zhang, and Y. He, "Optimized Bit Allocation for Learning-based Video Compression," in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, May 2022, pp. 1938–1942. doi: 10.1109/ISCAS48785.2022.9937681.
- [82] H. Wang, Z. Chen, and C. W. Chen, "Learned Video Compression via Heterogeneous Deformable Compensation Network," *IEEE Trans Multimedia*, vol. 26, pp. 1855–1866, Jul. 2024, doi: 10.1109/TMM.2023.3289763.
- [83] J. Shimizu, Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "HEVC Video Coding with Deep Learning Based Frame Interpolation," in *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, IEEE, Oct. 2020, pp. 433–434. doi: 10.1109/GCCE50665.2020.9291852.
- [84] R. Yano, Y. Liu, H. Watanabe, T. Suzuki, T. Chujoh, and T. Ikai, "A Study of Video Super-Resolution Method Using Video Coded Data as Training Data," in *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, IEEE, Oct. 2021, pp. 278–279. doi: 10.1109/GCCE53005.2021.9621883.
- [85] Y. Zhang, Z. Chen, and S. Liu, "A multi-stage spatio-temporal adaptive network for video super-resolution," *J Vis Commun Image Represent*, vol. 87, p. 103555, Aug. 2022, doi: 10.1016/j.jvcir.2022.103555.
- [86] D. Li, Z. Wang, and J. Yang, "Video super-resolution with inverse recurrent net and hybrid local fusion," *Neurocomputing*, vol. 489, pp. 40–51, Jun. 2022, doi: 10.1016/j.neucom.2022.03.019.
- [87] R. Lin, M. Wang, P. Zhang, S. Wang, and S. Kwong, "Multiple hypotheses based motion compensation for learned video compression," *Neurocomputing*, vol. 548, p. 126396, Sep. 2023, doi: 10.1016/j.neucom.2023.126396.

- [88] Y. Zhan, S. Xiong, X. He, B. Tang, and H. Chen, “A channel-wise contextual module for learned intra video compression,” *J Vis Commun Image Represent*, vol. 99, p. 104070, Mar. 2024, doi: 10.1016/j.jvcir.2024.104070.
- [89] C. Liu, “Fast HEVC inter-frame coding based on LSTM neural network technology,” *J Vis Commun Image Represent*, vol. 98, p. 104056, Feb. 2024, doi: 10.1016/j.jvcir.2024.104056.
- [90] Y. Zhang, S. Kwong, and S. Wang, “Machine learning based video coding optimizations: A survey,” *Inf Sci (N Y)*, vol. 506, pp. 395–423, Jan. 2020, doi: 10.1016/j.ins.2019.07.096.
- [91] F. Zhang, M. Afonso, and D. R. Bull, “ViSTRA2: Video coding using spatial resolution and effective bit depth adaptation,” *Signal Process Image Commun*, vol. 97, p. 116355, Sep. 2021, doi: 10.1016/j.image.2021.116355.
- [92] Y. Liu, S. Liu, Y. Wang, and H. Zhao, “Video coding and processing: A survey,” *Neurocomputing*, vol. 408, pp. 331–344, Sep. 2020, doi: 10.1016/j.neucom.2019.07.115.
- [93] M. Amna, W. Imen, S. F. Ezahra, and A. Mohamed, “Fast intra-coding unit partition decision in H.266/FVC based on deep learning,” *J Real Time Image Process*, vol. 17, no. 6, pp. 1971–1981, Dec. 2020, doi: 10.1007/S11554-020-00998-5/FIGURES/7.
- [94] C. S. Pawar and S. K. D. Sawarkar, “Multilevel ML Assistance to High Efficiency Video Coding,” *Advances in Intelligent Systems and Computing*, vol. 924, pp. 313–324, 2019, doi: 10.1007/978-981-13-6861-5_27/TABLES/2.
- [95] Y. Wang, X. Wu, X. He, C. Ren, and T. Zhang, “Deep Feature Fusion Network for Compressed Video Super-Resolution,” *Neural Process Lett*, vol. 54, no. 5, pp. 4427–4441, Oct. 2022, doi: 10.1007/S11063-022-10816-7/TABLES/3.
- [96] K. Da Ji and H. Hlavacs, “Deep Learning Based Video Compression,” *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 429 LNICST, pp. 127–141, 2022, doi: 10.1007/978-3-030-99188-3_8/FIGURES/7.
- [97] F. Mentzer, E. Agustsson, J. Ballé, D. Minnen, N. Johnston, and G. Toderici, “Neural Video Compression Using GANs for Detail Synthesis and Propagation,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13686 LNCS, pp. 562–578, 2022, doi: 10.1007/978-3-031-19809-0_32/TABLES/1.
- [98] N. Sigger, N. Al-Jawed, and T. Nguyen, “Spatial-Temporal Autoencoder with Attention Network for Video Compression,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13233 LNCS, pp. 290–300, 2022, doi: 10.1007/978-3-031-06433-3_25/TABLES/1.
- [99] C. Jia, W. Zhang, K. Lin, and S. Ma, “Towards Neural Video Compression: A Rate Distortion Modeling Perspective,” *Communications in Computer and Information Science*, vol. 1611 CCIS, pp. 26–40, 2022, doi: 10.1007/978-981-19-5096-4_3/FIGURES/10.
- [100] I. Valiandi, A. S. Panayides, E. Kyriacou, C. S. Pattichis, and M. S. Pattichis, “A Comparative Performance Assessment of Different Video Codecs,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14185 LNCS, pp. 265–275, 2023, doi: 10.1007/978-3-031-44240-7_26/FIGURES/5.
- [101] D. García-Lucas, G. Cebrián-Márquez, and P. Cuenca, “Rate-distortion/complexity analysis of HEVC, VVC and AV1 video codecs,” *Multimed Tools Appl*, vol. 79, no. 39–40, pp. 29621–29638, Oct. 2020, doi: 10.1007/S11042-020-09453-W/FIGURES/7.
- [102] R. Birman, Y. Segal, and O. Hadar, “Overview of Research in the field of Video Compression using Deep Neural Networks,” *Multimed Tools Appl*, vol. 79, no. 17–18, pp. 11699–11722, May 2020, doi: 10.1007/S11042-019-08572-3/FIGURES/21.
- [103] Y. Shi, Y. Ge, J. Wang, and J. Mao, “AlphaVC: High-Performance and Efficient Learned Video Compression,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13679 LNCS, pp. 616–631, 2022, doi: 10.1007/978-3-031-19800-7_36/TABLES/4.
- [104] A. Goliński, R. Pourreza, Y. Yang, G. Sautière, and T. S. Cohen, “Feedback Recurrent Autoencoder for Video Compression,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12625 LNCS, pp. 591–607, 2021, doi: 10.1007/978-3-030-69538-5_36/FIGURES/8.
- [105] J. Liu *et al.*, “Conditional Entropy Coding for Efficient Video Compression,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12362 LNCS, pp. 453–468, 2020, doi: 10.1007/978-3-030-58520-4_27/FIGURES/8.
- [106] H. Liu *et al.*, “Video super-resolution based on deep learning: a comprehensive survey,” *Artif Intell Rev*, vol. 55, no. 8, pp. 5981–6035, Dec. 2022, doi: 10.1007/S10462-022-10147-Y/TABLES/4.



- [107] T. Isobe, X. Jia, S. Gu, S. Li, S. Wang, and Q. Tian, "Video Super-Resolution with Recurrent Structure-Detail Network," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12357 LNCS, pp. 645–660, 2020, doi: 10.1007/978-3-030-58610-2_38/FIGURES/8.
- [108] L. Ge, D. Z. Jiang, and W. Bao, "3RE-Net: Joint Loss-REcovery and Super-REsolution Neural Network for REal-Time Video," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14471 LNAI, pp. 165–177, 2024, doi: 10.1007/978-981-99-8388-9_14/FIGURES/5.
- [109] J. Yin, S. Peng, J. Lin, D. Jiang, and C. Fang, "An Effective CNN-Based Super Resolution Method for Video Coding," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14233 LNCS, pp. 124–134, 2023, doi: 10.1007/978-3-031-43148-7_11/FIGURES/6.
- [110] M. M. Ho, G. He, Z. Wang, and J. Zhou, "Down-sampling based video coding with degradation-aware restoration-reconstruction deep neural network," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11961 LNCS, pp. 99–110, 2020, doi: 10.1007/978-3-030-37731-1_9/FIGURES/6.
- [111] X. Liu and Y. Wu, "An Efficient Compression Coding Method for Multimedia Video Data Based on CNN," *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 468 LNICST, pp. 123–136, 2023, doi: 10.1007/978-3-031-28787-9_10/TABLES/6.
- [112] Y. O. Sharrab, M. A. Alsmirat, M. A. H. Eljinini, and N. J. Sarhan, "iHELP: a model for instant learning of video coding in VR/AR real-time applications," *Multimed Tools Appl*, pp. 1–40, Mar. 2024, doi: 10.1007/S11042-024-18666-2/TABLES/3.
- [113] S. Ma *et al.*, "Overview of intelligent video coding: from model-based to learning-based approaches," *Visual Intelligence 2023 1:1*, vol. 1, no. 1, pp. 1–19, Aug. 2023, doi: 10.1007/S44267-023-00018-7.
- [114] H. K. Joy, M. R. Kounte, A. Chandrasekhar, and M. Paul, "Deep Learning Based Video Compression Techniques with Future Research Issues," *Wirel Pers Commun*, vol. 131, no. 4, pp. 2599–2625, Aug. 2023, doi: 10.1007/S11277-023-10558-2/FIGURES/13.
- [115] Q. Cao, D. Zhang, and C. Sun, "Quality Scalable Video Coding Based on Neural Representation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14554 LNCS, pp. 396–409, 2024, doi: 10.1007/978-3-031-53305-1_30/TABLES/5.
- [116] M. Mauri, T. Elli, G. Caviglia, G. Uboldi, and M. Azzi, "RAWGraphs: A visualisation platform to create open outputs," *ACM International Conference Proceeding Series*, vol. Part F131371, Sep. 2017, doi: 10.1145/3125571.3125585.
- [117] "Características | Flourish | Visualización de datos y narración de historias." Accessed: Jul. 14, 2024. [Online]. Available: <https://flourish.studio/features/>
- [118] "VOSviewer - Visualización de paisajes científicos." Accessed: Jul. 14, 2024. [Online]. Available: <https://www.vosviewer.com/>
- [119] N. J. van Eck and L. Waltman, "Visualizing Bibliometric Networks," *Measuring Scholarly Impact*, pp. 285–320, 2014, doi: 10.1007/978-3-319-10377-8_13.
- [120] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Comput Sci*, vol. 2, no. 6, pp. 1–20, Nov. 2021, doi: 10.1007/S42979-021-00815-1/FIGURES/11.
- [121] R. Arnosó Pérez, "Analysis and Evaluation of No-Reference Video Quality Assessment using Neural Networks Título: Analysis and Evaluation of No-Reference Video Quality Assessment using Neural Networks."
- [122] J. Zhu, H. Amirpour, R. Schatz, C. Timmerer, and P. Le Callet, "Enhancing Satisfied User Ratio (SUR) Prediction for VMAF Proxy through Video Quality Metrics," *2023 IEEE International Conference on Visual Communications and Image Processing, VCIP 2023*, 2023, doi: 10.1109/VCIP59821.2023.10402707.