Aprendizaje basado en prototipos para la identificación de fragmentos 3D en cerámicas íberas

Inmaculada Cubero León[1 [0009−0009−3757−4984]], José Manuel Fuertes García[1 [0000−0001−6624−4102]], Rafael Jesús Segura Sánchez[1 [0000−0002−3075−6963]], and Miguel Díaz Medina[1 [0000−0003−2577−323X]]

Departamento de Informática, Universidad de Jaén (Jaén, España)

Resumen La mayoría de materiales arqueológicos se encuentran afectados por factores ambientales o antrópicos, siendo su reconstrucción un proceso complejo de alta precisión cuya manipulación física debe minimizarse. En este escenario, la restitución se aborda desde la visión por computador, utilizando mayoritariamente representaciones 2D. En el dominio tridimensional, debido al coste computacional y el tiempo de adquisición, se suelen utilizar representaciones basadas en nubes de puntos. En este artículo, se aborda el problema de la identificación de fragmentos sobre cerámicas íberas a través de varias estrategias, donde se propone un primer enfoque, basado en el método de clasificación tradicional de muestras. Sin embargo, el costo y la escasez de *datasets* etiquetados requiere evolucionar hacia metodologías de entrenamiento auto/semi-supervisado. Atendiendo a esto, se presenta un segundo enfoque para la identificación de morfologías completas a partir de información poco significativa de la estructura original mediante aprendizaje por contraste. La experimentación se ha realizado sobre un conjunto de datos de fragmentos cerámicos formado a partir de cerámicas íberas originales con el fin de identificar, a partir de un fragmento dado, la tipología cerámica que presenta la máxima probabilidad de correspondencia.

Keywords: Computer Graphics · Geometric Deep Learning · Contrastive Learning

1. Introducción

Durante las excavaciones arqueológicas, es habitual encontrar materiales fracturados o dañados, siendo la cerámica uno de los hallazgos más frecuentes. Debido a su corta vida útil, estos objetos resultan fundamentales para analizar rasgos cronológicos y geográficos, ya que su forma y decoración experimentan cambios significativos en el tiempo y el espacio [6]. Este análisis constituye la base para la datación de los estratos arqueológicos y aporta datos valiosos sobre la producción local, las relaciones comerciales y el comportamiento de los consumidores [13,15]. Diversos estudios previos han utilizado perfiles 2D para la clasificación automática [4, 5, 19, 20] y la extracción de características [14, 23, 24, 33–35], empleando

desde técnicas tradicionales de procesamiento de imágenes hasta enfoques de aprendizaje profundo.

Dado que la gran mayoría de las piezas se recuperan fragmentadas, su reensamblaje manual es una tarea ardua y delicada que requiere una manipulación física que debe minimizarse por razones de conservación. Ante la aparición de sistemas de digitalización 3D asequibles, ha crecido el interés por el reensamblaje y la reconstrucción automática [7, 12, 32] y el análisis de fragmentos [38], aunque los trabajos existentes suelen limitarse a representaciones en 2D [26]. En trabajos previos, se ha propuesto un marco basado en una Red Generativa Adversaria con Autoencoder 3D (3D AE-GAN) aplicado a la reconstrucción y aumento de datos de cerámicas; en esta propuesta, dado un conjunto de fragmentos 3D conocidos, el modelo genera la pieza completa cuya geometría resulta más compatible con dichos fragmentos, proporcionando modelos virtuales con las mismas características geométricas que los reales. El conjunto de datos aumentado resultante será de gran utilidad para diseñar procedimientos automáticos que asistan a los expertos en la tarea de completado de las piezas [25]. Este enfoque fue abordado usando *grids* de vóxeles como formato de representación 3D, obteniendo resultados prometedores, si bien las resoluciones utilizadas eran necesariamente reducidas debido al compromiso que supone el uso de este formato en términos de almacenamiento en memoria.

En este trabajo se aborda el problema empleando una representación de los objetos 3D basada en nubes de puntos. Esta representación se ha consolidado como una de las más utilizadas en digitalización 3D por motivos claros: coste y tiempo de adquisición reducidos; es un formato simple que describe superficies mediante muestreo, sin requerir una topología compleja ni renderizados avanzados; permite ajustar la densidad de puntos según la necesidad de resolución y ofrecen alta fidelidad respecto a la escena real. Sin embargo, también presenta algunas desventajas: falta de estructura y orden, lo que dificulta su procesamiento; inexistencia de topología, al no haber información sobre la relación entre puntos vecinos, y la densidad varía según la distancia al sensor y el tipo de escáner, además de presentar habitualmente dispersión. A todo esto, se añade la probabilidad de ocurrencia del fenómeno de la oclusión, motivado por la imposibilidad de capturar objetos ocultos tras otros elementos. Y también, el coste computacional asociado al procesamiento de nubes masivas, del orden de varias decenas de millones de puntos, es significativo tanto en memoria como en tiempo de ejecución. En este sentido, las principales contribuciones de este trabajo son:

- Proponer una arquitectura basada en aprendizaje por contraste para la identificación de representaciones 3D morfológicamente similares que no requiera un conjunto de datos etiquetados.
- Realizar una comparativa entre un enfoque autosupervisado frente a paradigmas tradicionales supervisados que resuelven problemas de clasificación.
- Realizar una comparativa entre varios métodos de aprendizaje por contraste (supervisado y autosupervisado) considerando similitudes entre fragmento-objeto y fragmento-fragmento.

2. Trabajos previos

En los últimos años, la representación de formas 3D a partir de nubes de puntos se ha convertido en uno de los campos de investigación más destacables dentro de la visión por computador, motivada en todo o en parte por el auge reciente de los modelos de aprendizaje profundo en el contexto de la Inteligencia Artificial. Las técnicas de aprendizaje profundo, por su capacidad en el tratamiento de la información multimodal y captura de relaciones no lineales, son capaces de obtener resultados satisfactorios en el procesamiento de conjuntos de imágenes 2D; sin embargo, siguen existiendo grandes retos en el procesamiento de nubes de puntos debido a la ausencia de homogeneidad y correlación estructural. El problema de clasificación de nubes de puntos se ha convertido en la base del análisis de la geometría 3D; su objetivo es identificar una morfología a partir de la extracción de características sobre el conjunto de puntos original.

Los primeros enfoques no eran capaces de procesar directamente las nubes de puntos debido a la complejidad asociada a la falta de estructura natural, ausencia de una relación de orden y alta variabilidad [28], siendo necesario utilizar un conjunto de reglas para la extracción de características para, posteriormente, utilizar clasificadores basados en aprendizaje automático como las máquinas de vectores de soporte (SVM) [11], AdaBoost [17] o Random Forest [8] y predecir la etiqueta de la clase. Sin embargo, no tenían suficiente capacidad de adaptación y eran propensos al ruido [40].

El incremento de la potencia de cómputo y de las infraestructuras de procesamiento de datos ha impulsado la aplicación de técnicas de aprendizaje profundo en el análisis de nubes de puntos. En este contexto, Qi et al. [29] introdujeron la arquitectura PointNet, capaz de procesar nubes de puntos directamente. PointNet alinea la nube de puntos mediante una red de transformación (T-Net) para luego mapear cada punto a un espacio de mayor dimensión usando capas lineales. Finalmente, aplica una operación de *max-pooling*, garantizando la invarianza ante permutaciones y agregando las características individuales en un vector global que representa la geometría completa. Sin embargo, entre sus limitaciones se encuentra su incapacidad para extraer información del contexto local y abstraer relaciones geométricas basadas en vecindad.

Para resolver este problema, se propuso PointNet++ [30] que introduce una estructura de aprendizaje jerárquico. Esta arquitectura procesa la nube de puntos simulando operaciones convolucionales, permitiendo capturar tanto la estructura global como detalles geométricos locales. Cada nivel jerárquico se compone de una capa de abstracción de conjuntos (*Set Abstraction Layer*), subdividida en tres operaciones: muestreo, agrupamiento y extracción mediante PointNet. La capa de muestreo utiliza el algoritmo *Farthest Point Sampling* (FPS) para seleccionar un subconjunto de puntos que actuarán como centroides. La capa de agrupamiento delimita vecindarios alrededor de dichos centroides. Finalmente, una micro-red PointNet procesa cada vecindario local para codificar las relaciones topológicas de los puntos contiguos en vectores de características locales.

Más allá de la evolución de PointNet, la literatura alberga arquitecturas fundamentadas en el aprendizaje de características locales, tales como PointMLP

[22], PCNN [1], GAPointNet [2] o DGCNN [37]. Esta última, una Red Neuronal Convolucional de Grafos Dinámicos, destaca por el uso de grafos dirigidos y la introducción de un operador de convolución de aristas (EdgeConv). Este módulo reconstruye el grafo utilizando los vecinos más cercanos en el espacio de características en cada capa de la red, optimizando la captura de la topología local y garantizando la invarianza ante traslaciones y permutaciones de los puntos.

El análisis de nubes de puntos 3D ha experimentado una transición significativa desde los mecanismos de clasificación convencionales hacia paradigmas de aprendizaje más dinámicos y escalables. En este sentido, el aprendizaje autosupervisado permite la extracción de representaciones latentes robustas sin necesidad de etiquetado previo [9]. Bajo este paradigma, el aprendizaje por contraste (*Contrastive Learning*, CL) [10] ha emergido como una técnica fundamental que se basa en la premisa de que las proyecciones de una misma muestra sometida a distintas aumentaciones deben converger en el espacio de características, mientras que objetos de clases distintas deben segregarse. Una de las aplicaciones más influyentes de este enfoque es el modelo CLIP [31] que adopta dicho principio para alinear pares de imágenes con sus correspondientes descripciones textuales. Esta metodología ha sido trasladada al dominio 3D mediante PointCLIP [39], una arquitectura que aprovecha la red CLIP preentrenada para codificar nubes de puntos a través de su proyección en mapas de profundidad multivista.

La integración de mecanismos contrastivos en arquitecturas de AutoEncoders [27] ha demostrado ser eficaz ante tareas de clasificación y completado de formas, permitiendo la captura de características globales discriminativas. Algunos avances, como la técnica ToThePoint, proponen optimizar este proceso mediante la reutilización de características descartadas durante el *max-pooling* [18].

Además, en el contexto del aprendizaje *few-shot*, la identificación de geometrías 3D no observadas durante el entrenamiento se fundamenta en la capacidad de generalización de *Prototypical Networks* [36]. Durante la fase de inferencia, el modelo procesa un conjunto compuesto por muestras objetivo a través de un codificador para extraer sus correspondientes *embeddings* en un espacio latente. A partir del centroide de las representaciones vectoriales, el sistema clasifica nuevos elementos mediante métricas de distancia, donde la optimización por similitud semántica y geométrica permite el reconocimiento de formas complejas. El aprendizaje por contraste elimina la necesidad de un etiquetado masivo o reentrenamiento y dota al modelo de una mayor capacidad de generalización en geometrías 3D complejas.

3. Metodología

3.1. Preprocesamiento de datos

Para el problema a resolver, se requiere un conjunto de datos de fragmentos de objetos 3D que permita resolver el problema de la identificación morfológica ante partes incompletas de la geometría. En este sentido, con el propósito de garantizar la invarianza del modelo frente a transformaciones geométricas y

variaciones en la densidad de muestreo, se implementa una estrategia de aumento de datos (*data augmentation*) en cada iteración algorítmica. Dicho proceso aplica transformaciones afines estocásticas (rotación, escalado y traslación). Paralelamente, se simulan fenómenos de dispersión y oclusión mediante el descarte aleatorio de un porcentaje de la muestra (*dropout* de puntos), complementado con una permutación estocástica de sus índices para eliminar cualquier sesgo asociado al orden del vector de entrada. Finalmente, antes de procesar las nubes de puntos, se normalizan espacialmente, reescalando el objeto 3D para que quede centrado en el origen (0, 0, 0) y contenido en una esfera unitaria de radio 1.

3.2. Aprendizaje de representaciones de puntos

Para abordar el problema de la identificación de fragmentos 3D, resulta necesario transformar las nubes de puntos de entrada en descriptores compactos, altamente representativos y discriminativos. En este trabajo, se utiliza la arquitectura PointNet++ como *backbone* principal para la extracción de características geométricas locales y globales, una elección que se fundamenta en su capacidad nativa para procesar conjuntos de puntos, garantizando invarianza ante la variabilidad de la información de entrada.

La hipótesis inicial contempló la formulación del problema como un entorno de clasificación, empleando la arquitectura PointNet para realizar la correspondencia desde el espacio geométrico de una nube de puntos hacia una etiqueta de clase categórica. No obstante, este paradigma asume la existencia de un espacio de clases estático y acotado ($\mathcal{C}$ fijo), condición que resulta incoherente en escenarios reales como el campo de la exploración arqueológica. En dichos contextos, el conjunto de materiales es inherentemente dinámico y está sujeto al descubrimiento progresivo de tipologías morfológicas previamente no catalogadas.

Esta limitación metodológica impone un cambio de paradigma hacia un enfoque de aprendizaje de representaciones, permitiendo una mayor generalización de morfologías que se encuentran fuera de la distribución. En consecuencia, se transiciona de la clasificación categórica hacia el aprendizaje por contraste. Bajo este enfoque, la red no predice una etiqueta o clase, sino que proyecta las características geométricas del fragmento en un espacio latente de alta dimensionalidad, donde la similitud estructural se cuantifica mediante métricas de distancia.

Esta aproximación tiene la ventaja de no requerir necesariamente la ejecución de procesos de reentrenamiento o ajuste fino ante la inclusión de nuevas muestras. Para identificación de un nuevo modelo en el sistema, basta con realizar una única inferencia a través del *backbone* (previamente entrenado) para extraer su *embedding* característico. Posteriormente, el cálculo del mejor encaje geométrico se resuelve de manera computacionalmente eficiente mediante la evaluación de la similitud del vector generado frente a la base de datos de referencia (figura 1).

3.3. Aprendizaje por contraste de formas

El aprendizaje por contraste se refiere a un paradigma de aprendizaje autosupervisado cuyo principal objetivo es organizar dentro de un mismo espacio latente continuo un conjunto de muestras cuya proximidad refleja su similitud conceptual y estructural. En este contexto, el aprendizaje por contraste aplicado a formas 3D permite estructurar los modelos digitales atendiendo a su similitud morfológica, siendo capaz de aprender y codificar formas y curvaturas complejas.

Tal como se ha señalado, frente a los métodos de clasificación tradicionales que relacionan geometrías con etiquetas invariantes disjuntas, este modelo resuelve el problema mediante la optimización de distancias dentro de un espacio de *embeddings* de alta dimensionalidad.

El principio bajo el que se sustenta esta metodología consiste en aproximar aquellas representaciones geométricas que sean semánticamente similares, denominadas pares positivos y, paralelamente, alejar aquellas que sean conceptualmente diferentes, denominadas pares negativos. El procedimiento estándar para la configuración de los pares contrastivos consiste en tomar una muestra de referencia (en este caso, una nube de puntos) a la que se denomina ancla (x_a), sobre la cual se aplican un conjunto de transformaciones de aumento de datos para generar vistas alternativas de actuarán como pares positivos entre sí (x_t). El resto de instancias contenidas dentro del mismo lote se consideran directamente como pares negativos respecto a la muestra.

Los pares positivos y negativos se procesan por el codificador (en este caso, PointNet++), donde el *backbone* abstrae la geometría (características topológicas locales y globales) hacia un conjunto de vectores de características unificados ($\mathbf{z} \in \mathbb{R}^D$). Finalmente, el espacio latente se optimiza mediante la minimización de una función de pérdida, en este caso, NT-Xent [3]. Dentro del aprendizaje por contraste supervisado, el estándar de referencia para la optimización del espacio latente mediante etiquetas semánticas es la función de pérdida SupCon [16]. Estas funciones penalizan matemáticamente al modelo si los *embeddings* del par positivo están alejados o los pares negativos excesivamente próximos. El error se calcula evaluando la similitud o distancia entre los vectores resultantes.

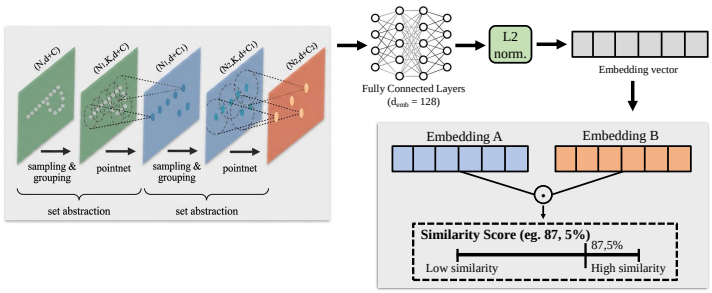


Figura 1. Modificación de la arquitectura PointNet++, para la comparativa entre fragmentos por similitud de *embedding*, basándose en la técnica *Contrastive Learning*.

4. Experimentación y resultados

4.1. Marco experimental

En este trabajo, se ha utilizado un *dataset* conformado por datos sintéticos de modelos 3D de fragmentos cerámicos construido a partir 1052 modelos digitales de cerámicas íberas originales. El conjunto de datos fue creado a partir de un fragmentador estocástico basado en la Cadena Discreta de Voronoi (DVC) [21]. Este *dataset* consta de aproximadamente 1M de nubes de puntos y mallas triangulares que supone un total de 450 GB de almacenamiento. El coste de procesar grandes cantidades de nubes de puntos afecta al tiempo de procesamiento y entrenamiento. En esta primera aproximación, se ha utilizado un subconjunto del *dataset* original (15 modelos 3D, un total de 14820 muestras de nubes de puntos de fragmentos), donde el 80 % se reserva para entrenamiento y el 20 % para test.

El marco experimental se estructura en dos enfoques metodológicos diferenciados: clasificación tradicional y aprendizaje por contraste; este último se evalúa bajo los paradigmas de aprendizaje supervisado y autosupervisado. Como resultado, se formulan las siguientes estrategias:

- **Estrategia 1 (CLS-SUP-FO).** Se considera el problema de pertenencia fragmento - objeto como un problema de clasificación convencional, considerando el modelo 3D (cerámica completa) como clase. Este enfoque trata de predecir a qué cerámica pertenece cada fragmento.
- **Estrategia 2 (CL-SUP-FF).** Aproximación contrastiva supervisada que define cuál es la mejor relación de afinidad entre fragmentos, aprovechando las etiquetas categóricas de pertenencia a una cerámica concreta. Se usan dos funciones de pérdida: SupCon (CL-SUP-FF-S) y NT-Xent (CL-SUP-FF-N).
- **Estrategia 3 (CL-SUP-FO).** Paradigma contrastivo supervisado que indica la cerámica más compatible dado un fragmento. Se utilizan etiquetas categóricas para indicar la relación entre fragmento y objeto completo. Se emplean dos funciones de pérdida: SupCon (CL-SUP-FO-S) y NT-Xent (CL-SUP-FO-N).
- **Estrategia 4 (CL-AUSUP-FF).** Resuelve la Estrategia 2 bajo un enfoque contrastivo autosupervisado mediante la generación de pares positivos y negativos sin etiquetas.

La Tabla 1 recoge los principales hiperparámetros de configuración de cada método, donde se indica el tamaño del lote, la tasa de aprendizaje (LR), la tasa de atenuación (DR), el número de épocas totales, el tipo de optimizador, la función de pérdida utilizada y el valor de temperatura para los enfoques contrastivos.

4.2. Resultados

Para abordar la tarea de identificación de fragmentos cerámicos, se evaluó el rendimiento de las estrategias propuestas. Los resultados de las métricas de evaluación se reflejan en la Tabla 2, donde se presenta el *accuracy* de cada

Tabla 1. Hiperparámetros utilizados para cada tipo de experimento.

Método	Tam. Lote	LR	DR	Época	Optimizador	Temperatura	Func. Pérdida
CLS-SUP-FO	64	0.001	0.0001	100	Adam	-	Cross-Entropy
CL-SUP-FF	64	0.001	0.0001	100	Adam	0.5	NT-Xent/SupCon
CL-SUP-FO	64	0.001	0.0001	100	Adam	0.5	NT-Xent/SupCon
CL-AUSUP-FO	64	0.001	0.0001	300	Adam	0.5	NT-Xent

método, evaluando si la predicción coincide con la máxima probabilidad (top-1) o se encuentra entre las cinco mejores (top-5). Asimismo, se indica la media de similitud entre pares positivos y negativos sobre los paradigmas contrastivos.

Tabla 2. Métricas de evaluación por método. Se detalla el *accuracy* (top-1 y top-5) en las fases de entrenamiento y test, junto a la media de similitud entre pares positivos y negativos.

Método	Accuracy		Accuracy (Top-5)		AVG Pos	AVG Neg	AVG Hard Neg
	Entrenamiento	Test	Entrenamiento	Test			
CLS-SUP-FO	0.934	0.876	**0.996**	0.988	-	-	-
CL-SUP-FF-N	0.983	**0.916**	0.985	0.949	0.986	**-0.078**	0.037
CL-SUP-FF-S	**0.985**	0.914	0.988	0.942	**0.99**	-0.076	**0.008**
CL-SUP-FO-N	0.919	0.762	0.957	0.824	0.916	-0.074	0.205
CL-SUP-FO-S	0.921	0.781	0.959	0.848	0.922	-0.072	0.195
CL-AUSUP-FF	0.82	0.905	0.99	**0.993**	0.961	-0.012	0.888

Como aproximación inicial, se propuso un primer enfoque (CLS-SUP-FO) utilizando la categorización de cada fragmento atendiendo a la cerámica a la que pertenece. Este planteamiento demostró una alta eficacia, registrando un *accuracy* en top-1 del 93.4 % en entrenamiento y 87.9 % en test. Sin embargo, esta aproximación require la presencia de clases fijas e invariantes, por lo que no permite adaptabilidad frente a nuevas muestras. Esto último motivó la transición hacia un paradigma más escalable, basado en la técnica de aprendizaje por contraste, que cuenta con la ventaja de no requerir necesariamente reentrenar el modelo ante la presencia de nuevas muestras (fragmentos). Bajo esta metodología, se consideraron varias estrategias: determinar la cerámica de mayor afinidad para el fragmento en cuestión (CL-SUP-FO) e identificar la instancia morfológicamente más similar en función de sus descriptores geométricos (CL-SUP-FF y CL-AUSUP-FF). La primera estrategia, que emplea las funciones de pérdida SupCon y NT-Xent, ofrece un alto rendimiento en entrenamiento, alcanzando un *accuracy* de $\sim 92\,\%$ en top-1 y $\sim 96\,\%$ en top-5. Sin embargo, ante muestras no vistas, este rendimiento disminuye en la fase de test: CL-SUP-FO-N registra un 76,2 % (top-1) y 82,4 % (top-5), mientras que CL-SUP-FO-S alcanza un 78,1 % (top-1) y 84,8 % (top-5). Respecto a las estrategias de similitud morfológica entre fragmentos (figura 2) se pueden observar resultados altamente satisfactorios, mostrando mejor rendimiento respecto a CL-SUP-FO. Destaca la versión autosupervisada que logra una precisión en test del 90.5 % (top-1) y 99.3 % (top-5).

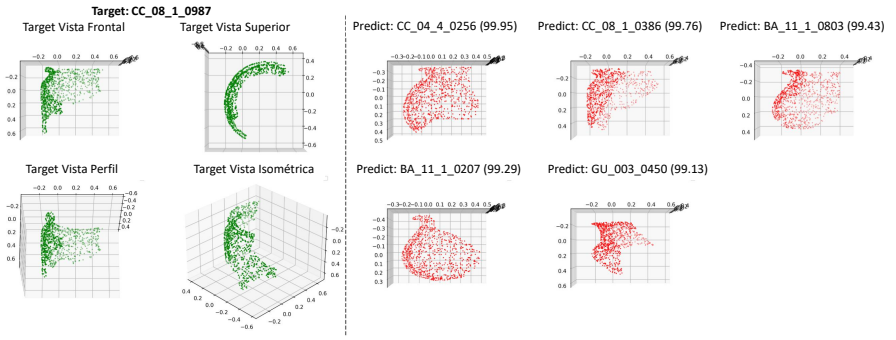


Figura 2. Predicción de la estrategia CL-AUSUP-FF: fragmento no visto por la red (izquierda) y fragmentos candidatos (derecha) ordenados por similitud. El segundo fragmento candidato es similar al objetivo y pertenece a la misma cerámica (CC_08_1).

Los resultados obtenidos con el *accuracy* están directamente relacionados con la capacidad del modelo de identificar las relaciones de similitud en el espacio latente. En este sentido, se cuenta con tres métricas adicionales que muestran el valor promedio de similitud entre pares positivos (AVG Pos), pares negativos (AVG Neg) y los valores con los que más se confunde el modelo (AVG Hard Neg). Como se puede observar en la Tabla 2, en todos los métodos se logra un valor elevado de similitid entre pares positivos y un valor bajo respecto a la similitud con pares negativos. En el paradigma autosupervisado, se puede observar cierta confusión del modelo en la identificación de fragmentos, ya que la única evidencia positiva es el par formado a partir de la aumentación de una muestra. No obstante, estos resultados demuestran que aplicar la técnica de aprendizaje por contraste autosupervisado sobre representaciones 3D permite abstraer las características geométricas y establecer fuertes correlaciones entre las muestras sin necesidad de utilizar etiquetas categóricas. Consecuentemente, se evita la complejidad inherente al etiquetado masivo de conjuntos de datos a gran escala y se ofrece un enfoque más optimizado en procesamiento.

5. Conclusiones y trabajo futuro

5.1. Conclusiones

Debido a las restricciones de almacenamiento y coste computacional, la mayoría de conjuntos de datos disponibles covergen hacia modalidades de baja dimensionalidad como representaciones 2D. Además, los *datasets* de modelos 3D presentan un gran inconveniente: ascienden rápidamente en memoria haciendo inviable procesar datos de calidad aceptable. Esto obliga a buscar otro tipo de representación con resoluciones más bajas, como nubes de puntos. Esto último se ve condicionado por la escasez de conjuntos de datos etiquetados, debido a la complejidad de categorizar cantidades masivas de datos. Esta limitación provo-

ca que se trascienda hacia enfoques menos restrictivos y más eficientes como el aprendizaje por contraste autosupervisado.

A partir de este paradigma, se pueden extraer características representativas de muestras sin necesidad de realizar una categorización previa ni requerir reentrenar la red ante nuevas muestras no identificadas previamente. Esta aproximación ofrece mejoras de rendimiento y optimización respecto a enfoques de clasificación tradicionales denotando resultados satisfactorios (sección 4.2).

5.2. Trabajo futuro

Esta prueba de concepto se ha realizado tomando como punto de partida un subconjunto de 15 modelos cerámicos; no obstante, en versiones posteriores, se pretende aplicar este enfoque sobre el *dataset* completo para determinar si la estrategia planteada es capaz de generalizar sobre un conjunto de datos significativamente más extenso. Tal como se ha presentado en la tabla 2, el enfoque autosupervisado fragmento - fragmento presenta muy buenos resultados; aún así, durante el proceso de entrenamiento se puede observar cierta confusión con otras muestras (AVG Hard Neg). Esto se debe a que en este enfoque sólo se considera par positivo a la versión aumentada de un mismo fragmento, siendo el resto pares negativos, indicando al modelo que un fragmento está relacionado consigo mismo, pero no con el resto. Sin embargo, dentro de un mismo lote, se pueden encontrar fragmentos de una misma cerámica o muy similares, introduciendo una contradicción implícita, ya que el sistema los procesa como si no tuvieran relación entre sí.

Atendiendo a esto, como trabajo futuro, se propone un enfoque de similitud jerárquica, estableciendo una similitud más fuerte entre los fragmentos resultantes de la aumentación de una muestra y una similitud más débil entre fragmentos distintos pertenecientes a la misma cerámica o geométricamente similares. Además, se pretende introducir una comparativa frente a otros modelos de referencia, ampliando y profundizando un marco experimental más completo. Junto a esto último, se plantea explorar y aprovechar modelos fundacionales pre-entrenados sobre los cuales aplicar *fine-tuning* para ajustarlos al problema de identificación de fragmentos 3D. Finalmente, se explorarán arquitecturas modernas para la extracción de características de nubes de puntos, como la arquitectura *transformer*, empleando el mecanismo de atención para identificar relaciones latentes entre pares de puntos.

Referencias

1. Atzmon, M., Maron, H., Lipman, Y.: Point convolutional neural networks by extension operators (2018)
2. Chen, C., Fragonara, L.Z., Tsourdos, A.: Gapointnet: Graph attention based point neural network for exploiting local feature of point cloud. Neurocomputing **438**, 122–132 (2021)
3. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.E.: A simple framework for contrastive learning of visual representations. ArXiv **abs/2002.05709** (2020)

4. Cintas, C., Speakman, S., Akinwande, V., Ogallo, W., Weldemariam, K., Sridharan, S., McFowland, E.: Detecting adversarial attacks via subset scanning of autoencoder activations and reconstruction error. In: IJCAI 2020. pp. 876–882 (2020)
5. Di Angelo, L., Di Stefano, P., Pane, C.: Automatic dimensional characterisation of pottery. Journal of Cultural Heritage **26**, 118–128 (2017)
6. Eslami, D., Di Angelo, L., Di Stefano, P., Pane, C.: Review of computer-based methods for archaeological ceramic sherds reconstruction. Virtual Archaeology Review **11**(23), 34–49 (2020)
7. Fragkos, Stergios, Tzimtzimis, Emanuel, Tzetzis, Dimitrios, Dodun, Oana, Kyratsis, Panagiotis: 3d laser scanning and digital restoration of an archaeological find. MATEC Web Conf. **178**, 03013 (2018)
8. Gan, Z., Zhong, L., Li, Y., Guan, H.: A random forest based method for urban object classification using lidar data and aerial imagery. In: 2015 23rd international conference on geoinformatics. pp. 1–4. IEEE (2015)
9. Goyal, P., Caron, M., Lefaudeux, B., Xu, M., Wang, P., Pai, V., Singh, M., Liptchinsky, V., Misra, I., Joulin, A., Bojanowski, P.: Self-supervised pretraining of visual features in the wild (2021)
10. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). vol. 2, pp. 1735–1742 (2006)
11. Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B.: Support vector machines. IEEE Intelligent Systems and their applications **13**(4), 18–28 (1998)
12. Kalasarinis, I., Koutsoudis, A.: Assisting pottery restoration procedures with digital technologies. International Journal of Computational Methods in Heritage Science (IJCMHS) **3**(1), 20–32 (2019)
13. Kampel, M., Sablatnig, R.: An automated pottery archival and reconstruction system. Journal of Visualization and Computer Animation **14**(3), 111–120 (2003)
14. Karasik, A., Smilansky, U.: Computerized morphological classification of ceramics. Journal of Archaeological Science **38**(10), 2644 – 2657 (2011)
15. Kashihara, K.: An intelligent computer assistance system for artifact restoration based on genetic algorithms with plane image features. International Journal of Computational Intelligence and Applications **16**(03), 1750021 (2017)
16. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 18661–18673. Curran Associates, Inc. (2020)
17. Kumar, R., Verma, R.: Classification algorithms for data mining: A survey. International Journal of Innovations in Engineering and Technology **1**(2), 7–14 (2012)
18. Li, X., Chen, J., Ouyang, J., Deng, H., Velipasalar, S., Wu, D.: Tothepoint: Efficient contrastive learning of 3d point clouds via recycling. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21781–21790 (2023)
19. Llamas, J., Lerones, P.M., Zalama, E., Gómez-García-Bermejo, J.: Applying deep learning techniques to cultural heritage images within the INCEPTION project. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 25–32 (2016)
20. Lucena, M., Fuertes, J.M., Martínez-Carrillo, A.L., Ruiz, A., Carrascosa, F.: Classification of archaeological pottery profiles using modal analysis. Multimedia Tools and Applications **76**(20), 21565–21577 (Oct 2017)
21. López, A., Rueda, A.J., Segura, R.J., Ogayar, C.J., Navarro, P., Fuertes, J.M.: Generating implicit object fragment datasets for machine learning. Computers & Graphics **125**, 104104 (2024)

22. Ma, X., Qin, C., You, H., Ran, H., Fu, Y.: Rethinking network design and local geometry in point cloud: A simple residual mlp framework (2022)
23. Mom, V.: Where did i see you before... a holistic method to compare and find archaeological artifacts. In: Decker, R., Lenz, H.J. (eds.) Advances in Data Analysis. pp. 671–680. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
24. Nautiyal, V., Kaushik, V.D., Pathak, V.K., Dhande, S., Nautiyal, S., Naithani, M., Juyal, S., Gupta, R.K., Vasisth, A.K., Verna, K.K., Singh, A.: Geometric modeling of indian archaeological pottery: A preliminary study. In: Clark, J., Hagemeister, E. (eds.) Exploring New Frontiers in Human Heritage. CAA2006. Computer Applications and Quantitative Methods in Archaeology. Fargo, United States (2006)
25. Navarro, P., Cintas, C., Lucena, M., Fuertes, J.M., Rueda, A., Segura, R., Ogayar-Anguita, C., González-José, R., Delrieux, C.: Iberianvoxel: automatic completion of iberian ceramics for cultural heritage studies. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. pp. 5833–5841 (2023)
26. Navarro, P., Cintas, C., Lucena, M., Fuertes, J.M., Segura, R., Delrieux, C., González-José, R.: Reconstruction of iberian ceramic potteries using generative adversarial networks. Scientific reports **12**(1), 1–11 (2022)
27. Nazir, D., Afzal, M.Z., Pagani, A., Liwicki, M., Stricker, D.: Contrastive learning for 3d point clouds classification and shape completion. Sensors **21** (11 2021)
28. Nguyen, A.T.L., Le, H.B.: 3D point cloud segmentation: A survey. 2013 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM) pp. 225–230 (2013)
29. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)
30. Qi, C.R., Yi, L., Su, H., Guibas, L.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017)
31. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision (2021)
32. Rasheed, N.A., Nordin, M.J.: A polynomial function in the automatic reconstruction of fragmented objects. J. Comput. Sci. **10**(11), 2339–2348 (2014)
33. Rice, P.M.: Pottery Analysis. University of Chicago Press, Chicago (1987)
34. Saragusti, I., Karasik, A., Sharon, I., Smilansky, U.: Quantitative analysis of shape attributes based on contours and section profiles in artifact analysis. Journal of Archaeological Science **32**(6), 841 – 853 (2005)
35. Shennan, S., Wilcock, J.: Shape and style variation in central german bell beakers. Science and Archaeology **15**, 17–31 (1975)
36. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. p. 4080–4090. Curran Associates Inc., Red Hook, NY, USA (2017)
37. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds (2019)
38. Wilczek, J., Monna, F., Navarro, N., Chateau-Smith, C.: A computer tool to identify best matches for pottery fragments. Journal of Archaeological Science: Reports **37**, 102891 (2021)
39. Zhang, R., Guo, Z., Zhang, W., Li, K., Miao, X., Cui, B., Qiao, Y., Gao, P., Li, H.: Pointclip: Point cloud understanding by clip (2021)
40. Zhou, L., Liu, Y., Zhang, P., Bai, X., Gu, L., Zhou, J., Yao, Y., Harada, T., Zheng, J., Hancock, E.R.: Information bottleneck and selective noise supervision for zero-shot learning. Machine Learning **112**, 2239 – 2261 (2022)