

# Deep Clustering: desempeño y técnicas recientes

Daniel José Ruda Lobato  
Maestría en Economía  
Universidad EAFIT  
Medellín, Colombia  
djrual@eafit.edu.co

Diego Andrés Valderrama Laverde  
Maestría en Ciencias de los Datos y Analítica  
Universidad EAFIT  
Medellín, Colombia  
davalderri@eafit.edu.co

Catalina Piedrahita Jaramillo  
Maestría en Ciencias de los Datos y Analítica  
Universidad EAFIT  
Medellín, Colombia  
cpiedrahij@eafit.edu.co

**Resumen**— El objetivo de este trabajo es realizar una revisión de literatura del *Deep Clustering* para tener un mejor entendimiento de su mejor desempeño en comparación con los algoritmos tradicionales de agrupación frente a datos de imágenes, y conocer las técnicas que han surgido para analizar estos datos. Para realizar eso, se utilizó la metodología SLR, que permite realizar una exploración sistemática de la literatura relevante en el tema. Los resultados de la revisión muestran que los métodos de agrupación profunda superan a las técnicas tradicionales de *clustering* y que se mantienen en una mejora continua.

**Palabras clave**— *deep clustering, imágenes, revisión sistemática de literatura*

## I. INTRODUCCIÓN

Los algoritmos de *Machine Learning* para la agrupación o clasificación han sido ampliamente utilizados en la literatura por su grado de exactitud. Sin embargo, los métodos convencionales de agrupamiento suelen tener un bajo desempeño con datos de alta dimensión, debido a la ineficiencia de sus medidas de similitud [1].

A pesar de que existen los extendidos métodos de reducción de dimensionalidad y transformación de descriptores, la complejidad de los datos sigue siendo un reto. Es por esto, que con el desarrollo del *Deep Learning*, las redes neuronales pueden utilizarse para transformar los datos de tal forma que sean más fáciles de utilizar al momento de realizar agrupaciones. A esta forma en la que se utiliza el *Deep Learning* en tareas de *clustering* se le conoce como *Deep Clustering*.

El objetivo de este trabajo es realizar una revisión de literatura del *Deep Clustering* para tener un mayor entendimiento sobre su mejor desempeño en la identificación de descriptores en datos de alta dimensión como las imágenes, en comparación con los algoritmos convencionales de agrupación, y conocer las técnicas de agrupación profunda que están a la vanguardia en la clasificación en imágenes. Para llevar a cabo esta investigación se utilizó la metodología *Systematic Literature Review* (SLR), que consiste en un protocolo que permite realizar una revisión de literatura de forma rigurosa basándose en preguntas de investigación específicas, palabras claves y filtros sistemáticos.

Los resultados de la revisión muestran que el uso del aprendizaje profundo en problemas de agrupación se ha extendido en diferentes áreas pasando de visión por computador hasta medicina, dermatología y agricultura. Los métodos de *Deep Clustering* muestran un mejor desempeño que los algoritmos tradicionales de agrupación e, incluso, están surgiendo nuevos algoritmos de agrupación profunda que están superando a las técnicas modernas de este mismo tipo de aprendizaje.

## II. MARCO TEÓRICO DEL PROBLEMA

Los algoritmos de agrupación se diseñan para tratar con bases de datos con amplio rango de aplicaciones en la vida real; sin embargo, en situaciones con bases de datos de alta dimensión y complejidad, como las imágenes, estos algoritmos pueden sufrir pérdidas en desempeño [1]. Para tratar estos problemas, se hace uso de *Deep Learning* para realizar tareas de agrupación, lo que se conoce como *Deep Clustering*.

Según [2], la alta calidad de los métodos de agrupación utilizando aprendizaje profundo sigue un principio universal: se aprende la representación de descriptores utilizando redes neuronales profundas y se emplea la representación de características como elemento de entrada para las técnicas de agrupación. La Fig. 1 muestra la metodología del *Deep Learning* en tareas de agrupación según [2].

La arquitectura de la red neuronal profunda suele convertir el elemento de entrada (*input*) en un mapa de características que es usado para la agrupación. Los diferentes tipos de arquitecturas utilizadas de redes neuronales pueden ser: *Multilayer Perceptron* (MLP), *Convolutional Neural Network* (CNN), *Deep Belief Network* (DBN), *Variational Autoencoder* (VAE) y *Generative Adversarial Network* (GAN).

Las redes neuronales profundas pueden consistir en una capa que es la capa de salida para datos de baja dimensión; o en varias capas donde se incluyen descriptores más sofisticados y se mejora el cálculo.

La pérdida de no agrupación se basa en el conjunto de restricción de las redes neuronales convolucionales profundas y la pérdida del *cluster assignment hardening* se basa en la medida de similitud usada para encontrar la distancia entre

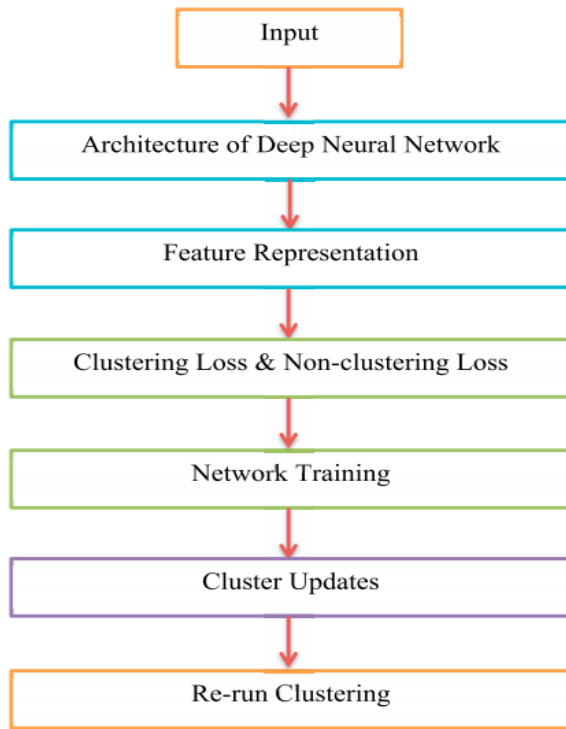


Fig. 1. La metodología de *Deep Learning* para enfoques de agrupación

puntos y centroides. La función de costo es el promedio ponderado de las pérdidas y la ponderación determina en qué grado se utiliza una función de pérdida para el entrenamiento de la red.

Durante el entrenamiento de la red, los centros y las asignaciones de los clústeres son actualizados. Las asignaciones se calculan usando probabilidades que tienen valores entre 0 y 1, valores que pueden ser tomados por la red, y usando un algoritmo de *back-propagation*, la arquitectura de la red se optimiza. Después de entrenar la red con la base de datos de entrenamiento, el algoritmo de agrupación puede ser ejecutado nuevamente usando la base de datos de prueba. Asimismo, se puede utilizar una base de datos similar para probar el desempeño de la red.

### III. METODOLOGÍA

Para realizar la revisión de literatura se utilizó la metodología de SLR, que permite identificar de forma rigurosa la literatura relacionada con preguntas de investigación específicas. Para identificar las preguntas de investigación se construyó un árbol de problemas (véase Fig. 2). El problema principal que se abordó consiste en el bajo desempeño que tienen los métodos tradicionales de agrupación (*clustering*) para identificar descriptores o características importantes en imágenes.

Este problema surge debido a que estos algoritmos no realizan un buen trabajo con datos de alta dimensión y

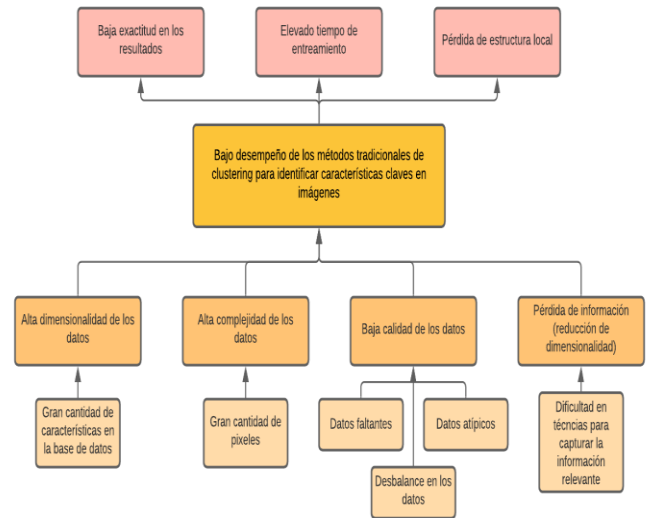


Fig. 2. Árbol del problema

complejidad, provocando bajos puntajes de exactitud en las pruebas y un elevado tiempo de entrenamiento. Sin embargo, los algoritmos de *Deep Clustering*, es decir, algoritmos de agrupación utilizando *Deep Learning*, están tomando ventaja en este sentido. Teniendo esto en cuenta, se establecieron las siguientes preguntas de búsqueda: ¿en qué medidas de desempeño el *Deep Clustering* supera al *clustering* tradicional en la clasificación de imágenes? ¿Qué técnicas de *Deep Clustering* están siendo utilizadas en la literatura para trabajar con imágenes?

Se tendrán en cuenta para la revisión de literatura los siguientes criterios de inclusión: 1) la revisión se realizará solamente sobre artículos científicos, 2) se tomarán en cuenta los artículos publicados en revistas después del 2015 y, 3) se considerarán solamente los artículos en inglés. Se excluirán los artículos no digitales y que no estén en su versión completa.

Se utilizó el servicio en línea de información científica *Web of Science* para identificar la literatura relevante para cada pregunta de investigación y se realizó la búsqueda de esa literatura en la base de datos bibliográfica *ScienceDirect*.

Para realizar las cadenas de búsqueda se identificaron los siguientes términos clave a partir de las preguntas de investigación:

A1 deep clustering

B1 images

B2 features

C1 performance

C2 accuracy

D1 methods

D2 techniques

Por lo tanto, cada pregunta de investigación tendrá una cadena de búsqueda con la que se identificará la literatura relevante relacionada en *Web of Science* y en *ScienceDirect*.

Pregunta 1: ¿En qué medidas de desempeño el *Deep Clustering* supera a los algoritmos de *clustering* tradicionales en la clasificación de imágenes?

A1 AND (B1 OR B2) AND (C1 OR C2)

Pregunta 2: ¿Qué técnicas de *Deep Clustering* están siendo utilizadas en la literatura para trabajar con imágenes?

A1 AND (B1 OR B2) AND (D1 AND D2)

Después de utilizar las cadenas de búsqueda en las bases de datos se procedió a realizar una revisión de los títulos y *abstracts* para descartar estudios que no vayan de acuerdo a los siguientes criterios de calidad: 1) que propongan una técnica de *Deep Clustering*, 2) que sean aplicados a datos de imágenes, 3) que realicen comparaciones con otras técnicas y, 4) que utilicen medidas para validar sus resultados.

En la sección siguiente se realiza la revisión de la literatura relevante identificada para cada pregunta de investigación.

#### IV. ANÁLISIS DE RESULTADOS

En esta sección se presentarán los resultados de la revisión de literatura utilizando la metodología SLR.

Pregunta 1: ¿En qué medidas de desempeño el *Deep Clustering* supera al *Machine Learning* tradicional en la clasificación de imágenes?

En la revisión de bibliografía para responder la pregunta se encontró un análisis estadístico que buscaba medir la relación entre la precisión de la agrupación y la calidad de la reconstrucción, lo cual es de interés en el campo de análisis de imágenes ya que en el proceso de compresión y descompresión se presenta una alta pérdida de información que disminuye el rendimiento del agrupamiento. Usualmente la dimensionalidad de los datos se reduce a través de transformaciones lineales y no lineales, como el análisis de componentes principales (PCA) y los métodos spectral/kernel, pero realizar *clustering* después de estas transformaciones resulta en pobres rendimientos [3] con el *Deep Clustering* es posible lograr mejores niveles de precisión, en especial por la calidad que tienen las redes neuronales de realizar una buena extracción de características sin supervisión, sobre todo en conjuntos de datos ricos.

Para realizar la extracción de las características se utiliza comúnmente los autocodificadores, los cuales son un tipo de red neuronal artificial que sirve para aprender descriptores de los datos de manera eficiente y sin supervisión, el objetivo del autocodificador es aprender una codificación de los datos para lograr una reducción de dimensionalidad, entrenando la red para ignorar el "ruido". En el estudio [3] se referencian ejemplos del alto potencial que se puede lograr utilizando las *Variational Autoencoders* (VAE) con diferentes conjuntos de datos, y utilizando diferentes arquitecturas de redes neuronales para extraer características de las imágenes, como por ejemplo el realizado por [4] en el que se logra dada una sola imagen de entrada, que el modelo genere nuevas

imágenes del mismo objeto con variaciones en la pose y la iluminación.

Con el objetivo entonces de lograr una precisión en la agrupación por medio de las características que se extraen de las imágenes y además una mejora en la calidad en la reconstrucción, los autores proponen una nueva función de costo para los VAE, los principales indicadores que se tuvieron en cuenta para este diseño fueron:

- Error cuadrado medio (MSE): Toma el promedio de las diferencias cuadradas entre los valores de píxeles de las imágenes de "referencia" de entrada y las imágenes reconstruidas "comprimidas o distorsionadas".
- Índice de calidad de imagen universal (UQI): Es un índice de calidad de imagen alternativo para modelar las imágenes reconstruidas mediante la incorporación de medidas de calidad de percepción, como la pérdida de correlación, distorsión de luminancia y contraste.
- Índice de fidelidad de la información visual (VIF). Es otro enfoque de evaluación de la calidad de referencia completa para medir la pérdida de información en una imagen reconstruida de la imagen de referencia que tiene una "calidad" perfecta percibida. Se basa en modelar las características del sistema visual humano (HVS) como un problema de fidelidad de la información. La idea detrás de VIF es simplemente medir la pérdida de información de la imagen (es decir, la incertidumbre que HVS agrega a la imagen de referencia) y cómo se relaciona con la calidad visual percibida.
- Índice de similaridad estructural (SSIM): Se define combinando los tres aspectos de la pérdida de información: las variaciones en la iluminación (luminancia), los cambios de brillo (contraste) y las comparaciones de estructura (distorsión de correlación).

Con la nueva función de costo los investigadores realizan ahora una prueba de *clustering* utilizando para ello el algoritmo estándar de *k-means*. Las imágenes pasan por el marco de VAE que aprende de forma supervisada cómo calcular la mejor representación de estas para el espacio latente, y esta codificación se utiliza para realizar la agrupación y lograr una mejor precisión. Se realizan diferentes experimentos cambiando el tamaño del espacio latente y de los clústeres. En la Fig. 3 se aprecia un esquema con la idea general de este trabajo.

Para evaluar el resultado de los agrupamientos se utilizan las métricas de precisión de agrupación no supervisada (ACC) e información mutua normalizada (NMI). La precisión se mide después de que se completa el entrenamiento VAE y el modelo aprende sus parámetros óptimos y las variables latentes. En la Fig. 4 se aprecia el marco utilizado en el estudio para realizar el análisis comparativo.

Aunque el *k-means* es un método de *Machine Learning* tradicional, en este estudio podemos apreciar como el éxito del proceso de *clustering* se centra en la exitosa extracción de características de la VAE y a su vez como soportan una exitosa

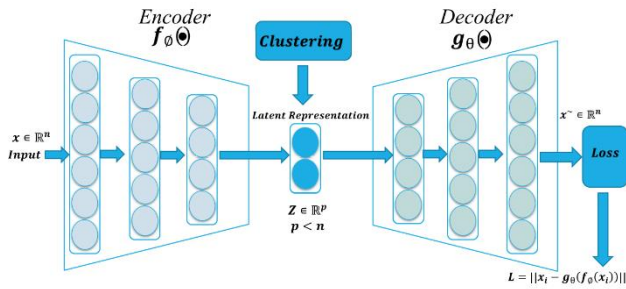


Fig. 3. Esquema simple del autocodificador variacional basado en agrupamiento [3]

reconstrucción de la imagen, conclusión que se fundamenta en los resultados obtenidos al realizar las métricas sobre las agrupaciones y las imágenes resultantes.

Con este análisis se demuestra una correlación positiva entre la precisión del agrupamiento que se obtiene por *k-means* en el cuello de botella de VAE y las métricas de calidad de reconstrucción de imágenes, SSIM, VIF y UQI, esto independientemente del tamaño del espacio latente. Se concluye también que el aumento de la dimensión del espacio latente no aumenta constantemente la precisión de la agrupación del *k-means*, ni mejora la calidad de reconstrucción de la imagen.

Con este trabajo una de las más importantes conclusiones de los investigadores fue las implicaciones futuras sobre la forma en que se pueden entrenar las topologías de *Deep Clustering* para una mayor precisión. La mayor ventaja es la reducción del tiempo de entrenamiento, especialmente cuando el parámetro de hiperespacio que se debe ajustar para un entrenamiento adecuado es lo suficientemente grande. Por ejemplo, la correlación demostrada entre la calidad de la imagen de reconstrucción y la precisión de la agrupación podría usarse para eliminar la necesidad de ejecutar el *back-end* de *k-means* cada vez que se realiza un nuevo barrido de hiperparámetros. Esto es específicamente más importante para

aplicaciones ricas en datos, como videos, donde cada fotograma consiste en una imagen estática que se puede usar para agrupar y, por lo tanto, una calidad de reconstrucción de imagen promedio calculada en todos los fotogramas se puede usar como indicador del rendimiento de agrupamiento final [3].

En general una de las más importantes características del *Deep Clustering* está relacionado con la selección de los descriptores, lo que hace que metodologías relacionadas a este sean usadas principalmente en problemas con clasificación no supervisada, por lo que para evidenciar el mejor desempeño del *Deep Clustering* con respecto a los modelos de *Machine Learning* tradicional se muestran también los resultados obtenidos en el artículo DIC: *Deep Image Clustering for Unsupervised Image Segmentation* [5], ya que aquí se puede apreciar como los investigadores realizan un agrupamiento con aprendizaje profundo y también aprovechan el potencial de las redes neuronales para la extracción de características.

En este caso se puede apreciar como los investigadores proponen un marco para realizar agrupamiento con aprendizaje profundo, para la segmentación de imágenes no supervisada. Realizan una subred para extracción de características y una *para Deep Clustering*, luego realizan una pérdida de refinamiento iterativo guiada por superpíxeles y finalmente un protocolo de entrenamiento de sobreajuste para optimizar los parámetros de la red de un extremo a otro.

Las métricas utilizadas para evaluar los resultados de la segmentación fueron

- Índice Rand Probabilístico (PRI): El índice Rand calcula una medida de similitud entre dos agrupamientos considerando todos los pares de muestras y contando los pares que se asignan en el mismo grupo o en grupos diferentes en los agrupamientos predichos y verdaderos.

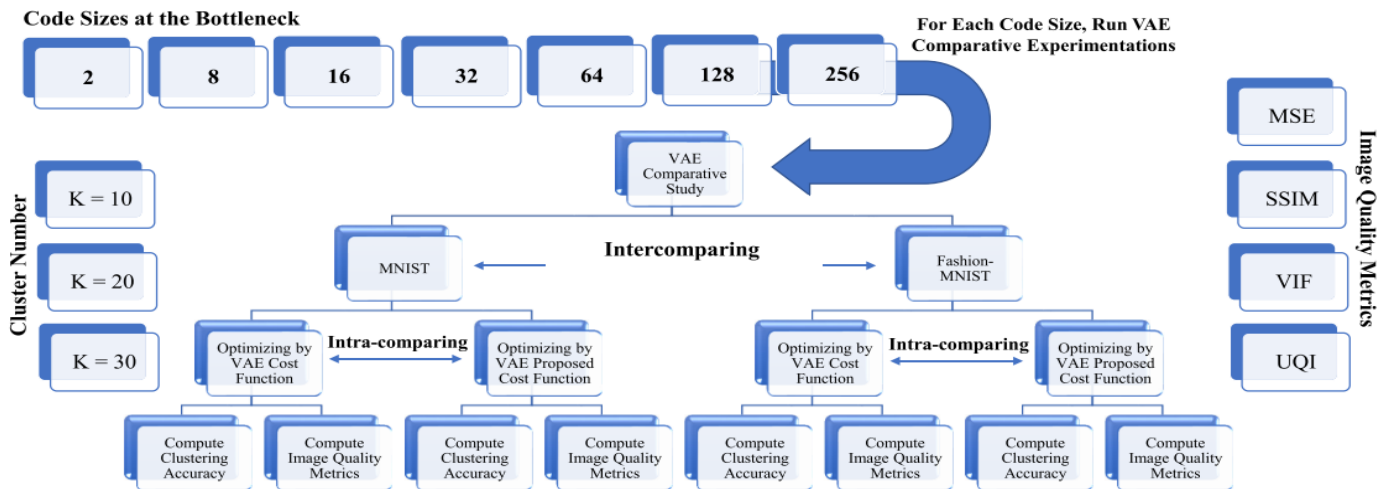


Fig. 4. Marco de trabajo análisis comparativo [3]

- Variación de Información (VoI): Variación de la información, que mide la distancia de la información que comparten dos agrupaciones
- Error de Coherencia Global (GCE): Mide hasta qué punto una segmentación puede verse como un refinamiento de la otra. Las segmentaciones que se relacionan de esta manera se consideran consistentes, ya que podrían representar la misma imagen natural segmentada a diferentes escalas.
- Error de Desplazamiento del Limite (BDE): Mide el error de desplazamiento promedio de los píxeles de límite entre dos imágenes segmentadas. En particular, define el error de un píxel de límite como la distancia entre el píxel y el píxel más cercano en la otra imagen de límite.

En la Tabla 1 se muestra un resumen de la investigación que tiene el comparativo entre las diferentes técnicas utilizadas y el *Deep Imagen Clustering* (DIC).

En el marco presentando por los investigadores para realizar la segmentación de imágenes sin supervisión con base a una red neuronal, se puede concluir al evaluar los resultados con las métricas de PIR, VoI, GCE y BDE, cómo con esta metodología se logran resultados satisfactorios. El proceso realizado por los investigadores fue dividir el espacio de imágenes en diferentes agrupaciones y a su vez realizar una actualización de los centros de la agrupación de forma iterativa, también se resalta la contribución de la incorporación de superpíxeles de bajo nivel en DIC mediante el diseño de una pérdida de refinamiento iterativa para optimizar los parámetros de la red de un extremo a otro, esto ayuda a que los superpíxeles puedan servir de señales de agrupación para la codificación compleja de patrones de imágenes. Además, DIC se utiliza para generar superpíxeles aprendiendo de patrones de imagen locales de una manera sobreajustada. En la Fig. 5 se aprecia cual fue el esquema general del marco de trabajo propuesto en esta investigación.

	PRI	VoI	GCE	BDE
Ncut	0.7242	2.9061	0.2232	17.15
Mean Shift	0.7958	1.9725	0.1888	14.41
FH	0.7139	3.3949	0.1746	16.67
JSEG	0.7756	1.8217	0.1989	14.40
MNcut	0.7559	2.4401	0.1925	15.10
NTP	0.7521	2.4954	0.2373	16.30
KM	0.76	2.4	X	X
SDTV	0.7758	1.8165	0.1768	16.24
ASC	0.7738	2.0292	0.2225	16.09
gPb-Hoiem	0.724	3.149	NA	14.795
TBES	0.80	1.76	NA	NA
W-Net	0.81	1.71	NA	NA
gPb-owt-ucm	0.81	<b>1.68</b>	NA	NA
CCP	0.8014	2.4723	<b>0.1270</b>	11.29
MLSS	0.8146	1.8545	0.1809	12.21
SAS	0.8319	1.6849	0.1779	11.29
DIC (proposed)	<b>0.8407</b>	1.7491	0.1390	<b>10.18</b>

Tabla 1. Evaluación del desempeño del método propuesto en comparación con otros últimos avances en la base de datos de segmentación 300 de Berkeley (BDS300) [5]

Con este revisión podemos evidenciar como el *Deep Clustering* es una gran herramienta para realizar la identificación de características claves en las imágenes, lo que es una importante variable de entrada para obtener mejores resultados en los proceso de agrupación, es de destacar por ejemplo los resultados que se logran con las VAE donde se realizar una reducción de dimensionalidad para facilitar el proceso de *clustering* y luego recuperación de información que mantiene una gran similaridad con la original, esto permite que el aprendizaje profundo sea una gran herramienta de soporte para acompañar cualquier otra técnica de *clustering*, aunque con el *Deep Clustering* en si se obtiene también de forma exitosa agrupaciones de imágenes como se puedo evidenciar con los resultados de [5]

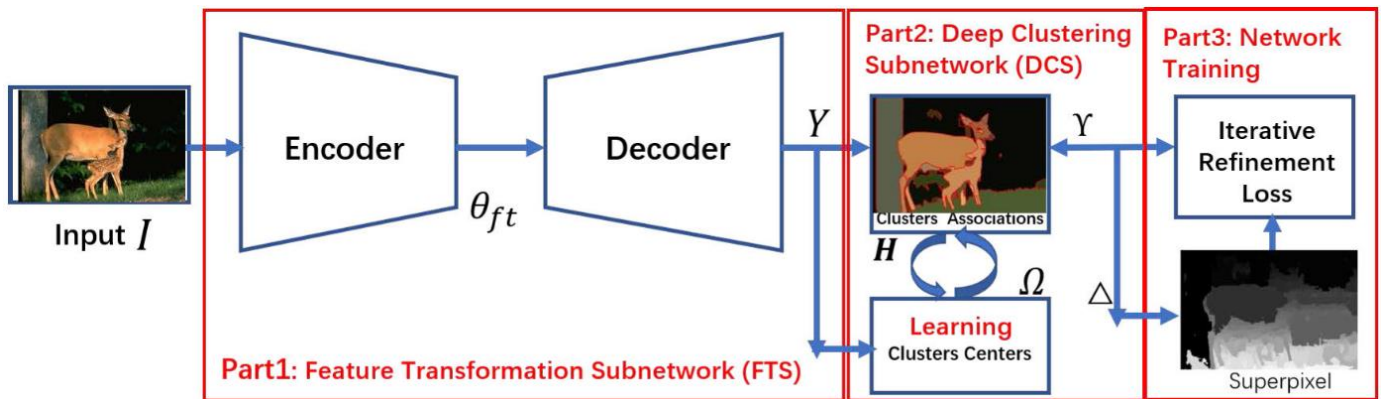


Fig. 5. La ilustración del marco DIC propuesto para la segmentación de imágenes sin supervisión

Pregunta 2: ¿Qué técnicas de *Deep Clustering* están siendo utilizadas en la literatura para trabajar con imágenes?

Cada vez más surgen nuevos métodos de agrupación profunda para la clasificación de datos de alta dimensión como lo son las imágenes, y con muchas aplicaciones interesantes en diversos campos.

En el campo de la agricultura se ha investigado mucho utilizando técnicas como las Redes Neuronales Convolucionales (CNN); sin embargo, estos trabajos todavía dependen del etiquetado de datos manual, lo cual involucra mucho tiempo. Por otra parte, el *Deep Learning* no supervisado no depende del etiquetado y puede ser el futuro en el área.

En [6] se prueban dos algoritmos de *Deep Clustering* no supervisados recientes: el *Joint Unsupervised Learning of Deep Representations and Image Clusters* (JULE) y el *Deep Clustering for Unsupervised Learning of Visual Feature* (*DeepCluster*) utilizando dos bases de datos públicos para la clasificación de diferentes tipos de hierbas. La primera base de datos se obtuvo en una plantación de soja en Brasil y discrimina la hierba entre pasto y hierba de hojas anchas (base *Grass-Broadleaf*). La segunda base de datos consiste en 17509 imágenes etiquetadas de ocho especies de hierbas nacionales nativas de Australia (base *DeepWeed*).

El desempeño de agrupación no supervisada fue evaluado utilizando las métricas del *Normalized Mutual Information* (NMI) y de *Unsupervised Clustering Accuracy* (UCA), y se analizaron los efectos de técnicas como el aumento de datos y el aprendizaje por transferencia para mejorar la calidad del agrupamiento lo cual establece una discusión que puede ser útil para el *Deep Clustering* no supervisado en general. También propusieron el uso de etiquetado de datos semi-automático, que reduce grandemente los costos de etiquetado manual. Este enfoque alcanza 97 % de exactitud en discriminar el pasto de las hojas anchas mientras reduce el número de anotaciones manuales en 100 veces.

JULE es un algoritmo de *Deep Clustering* no supervisado que usa un agrupamiento aglomerativo junto con CNN para generar clústeres de imágenes. La idea principal detrás del agrupamiento aglomerativo es unir dos clústeres en cada paso usando una medida de afinidad definida hasta que se alcance un criterio de parada; por ejemplo, el número de clústeres es igual a un número  $n$  definido por el usuario. Esta tarea es ejecutada a través de un marco recurrente que usa las representaciones de las CNN previas para generar las etiquetas de los clústeres para cada imagen y actualizar los pesos de este CNN usando estas etiquetas de clústeres. La función de pérdida se calcula usando la medida de afinidad entre los clústeres.

Como en los primeros pasos, las etiquetas generadas por los pesos aleatorios del CNN no son muy fiables, el algoritmo inicialmente realiza sobre-agrupamiento, que se unen mientras las mejores representaciones CNN se aprenden.

*DeepCluster* es un enfoque de agrupamiento escalable para el aprendizaje no supervisado de CNN que alternan entre agrupamiento de descriptores de imágenes y actualización de pesos de la CNN prediciendo las asignaciones de clústeres.

Para este trabajo, el algoritmo alterna entre agrupar las características con algoritmos estándar de agrupación para producir pseudo-etiquetas y actualizar los parámetros de la CNN prediciendo estas pseudo-etiquetas.

La principal idea detrás del *DeepCluster* es explotar el poder discriminativo de una CNN. El resultado del CNN se agrupa y las asignaciones de clústeres subsecuente se usan para optimizar la pérdida de la red. Este enfoque de *Deep Clustering* aprende iterativamente las características y las agrupa. Sin embargo, como este tipo de procedimiento alterno es propenso a soluciones triviales tales como clústeres vacíos y parametrización trivial, el método implementa algunas técnicas para tratar de evitar la degradación del aprendizaje de la red.

Analizando los dos algoritmos de agrupación evaluados, JULE tuvo un peor desempeño que *DeepCluster* según ACC y NMI. Este resultado se esperaba ya que JULE fue implementado tomando un número menor de clústeres y fue probado en la base de datos más simple como *Grass-Broadleaf*. En la compleja base de datos *DeepWeeds*, los resultados fueron más cercanos a una agrupación aleatoria de imágenes. Por otro lado, *DeepCluster* logró resultados interesantes en ambas bases de datos utilizando un gran número de clústeres, pero estos resultados empeoraban cuando disminuía el número de clústeres.

Otro campo en donde se utilizan algoritmos de *Deep Learning* para agrupación en imágenes es en el de la dermatología. La segmentación automática de lesiones de piel en imágenes de dermatoscopia para reconocer melanomas de piel a través de sistemas de diagnóstico asistidos por ordenador es una tarea desafiante.

En [7] proponen una nueva metodología de segmentación vía *Full Resolution Convolutional Networks* (FrCN). Este método aprende directamente las características con resolución completa de cada píxel individual de los datos de entrenamiento sin la necesidad de operaciones de procesamiento tales como la eliminación de artefactos, el ajuste de contraste bajo y mejora de los límites de lesiones de piel segmentadas.

El método FrCN fue probado usando dos bases de datos disponibles al público: la base de datos del Simposio Internacional en Imágenes Biomédicas (ISBI en inglés) del 2017 del IEEE, y la base de datos PH2. Para evaluar el método, se compararon los desempeños en segmentación con los enfoques más recientes de segmentación con *Deep Learning* tales como el *Full Convolutional Network* (FCN), *U-Net* y *SegNet*.

En el método FrCN las características de completa resolución de los píxeles individuales de entrada son reservados eliminando todas las capas de submuestreo de la arquitectura. Cada píxel puede ser representado como una muestra de entrenamiento

La primera parte consiste en capas convolucionales que son responsables de extraer los descriptores de las imágenes de entrada usando varios tipos de filtros. En este método no existen capas de submuestreo que puedan reducir la resolución de las características espaciales de la imagen de entrada. En



este caso, cada píxel puede ser representado como una muestra de entrenamiento. La última capa del FrCN es alimentada por una regresión logística multinomial conocida como el clasificador *softmax*, que usa los descriptores extraídos de cada píxel para clasificarlo como lesión o tejido, lo que produce un mapa de predicciones para cada píxel.

En la base de datos de ISBI 2017, el método FrCN sobrepasó a FCN, U-Net y *SegNet* con los índices de precisión general, Dice y Jaccard de 94.03 %, 87.08 % y 77.11 %, respectivamente; además, el AUC general fue mayor en el caso del FrCN. En la base de datos PH2, FrCN supera los demás métodos. La exactitud en la segmentación de FCN, U-Net y *SegNet* y FrCN fueron 92.82 %, 92.55 %, 93.36 % y 95.08 %, respectivamente. En el caso de AUC, FrCN se desempeña ligeramente mejor en los casos benignos, mientras demostró una mejora significativa en los casos de melanoma.

Por todo lo anterior, el uso de resoluciones espaciales completas de la imagen de entrada puede permitir aprender mejor descriptores específicos y prominentes, conllevando a una mejora en el desempeño de la segmentación.

En la medicina, una importante tarea es diagnosticar subtipos de enfermedades bastantes distintas para un tratamiento efectivo. En [8] se propone un nuevo modelo de *Deep Learning* y *clustering* que combina *Deep Belief Network* (DBN) y *Fuzzy C-Means* (FCM), llamado *Unsupervised Deep Fuzzy C-Means clustering Network* (UDFCMN) para agrupar pacientes con cáncer de pulmón a partir de imágenes de tomografías computarizadas (CT en inglés) de pulmones.

Después de su procesamiento, primero las imágenes son codificadas en múltiples capas de variables escondidas para extraer las características jerárquicas y la distribución de las características, y formar representaciones de alto nivel. Para solucionar los problemas de la homogeneización de las características en la DBN se introdujo el *Winner-Take-All* DBN (WTADBN). Se usa FCM para producir las etiquetas iniciales de los clústeres con las nuevas representaciones aprendidas por una WTARBM apilada. De esta forma, las etiquetas de los clústeres generadas por FCM se usan para los ajustes del DBN como etiquetas reales. Así, la agrupación no supervisada de imágenes y el proceso de estratificación de pacientes se completa por iteración cruzada.

El algoritmo de agrupación fue probado para experimentar tanto en una base de datos pública de internet como una base de datos privada de un hospital cooperativo. Los resultados muestran que el algoritmo UDFCMN supera a los métodos de clasificación no supervisados de última generación. Estos resultados también indican que el algoritmo puede tener aplicaciones prácticas en los estudios patogénesis del cáncer de pulmón y provee una útil guía para terapia de cáncer personalizada.

La segmentación no supervisada es una técnica esencial previa al procesamiento en muchas tareas de visión por computador; sin embargo, las técnicas recientes de segmentación no supervisada son sensibles a los parámetros tales como el número de segmentación o la complejidad de alto entrenamiento e inferencia. En [5] se propone un marco de segmentación no supervisada basada en un nuevo modelo

de agrupación profunda de imágenes (*Deep Image Clustering*, DIC), que consiste en una subred de transformación de características (*Feature Transformation Subnetwork*, FTS) y una subred de *Deep Clustering* entrenable (DCS) para la agrupación de imágenes no supervisada.

La FTS se basa en una arquitectura de red simple y capaz, y el DCS puede asignar píxeles con números de clústeres diferentes actualizando las asociaciones y los centros de los clústeres iterativamente. Además, una pérdida del refinamiento iterativo guiado por super píxeles se diseñó para optimizar los parámetros de DIC en una manera de sobreajuste. Se utilizó la base de datos *Berkley Segmentation* y los resultados muestran que DCS es más efectivo en agregar los descriptores durante el proceso de *clustering*. Asimismo, DIC ha probado ser menos sensible a los parámetros de segmentación variantes y de costos de computaciones menores, y puede lograr significativamente mejor desempeño en segmentación comparado con otras técnicas de nueva generación.

En [9] proponen a *Two-Stage Deep Density-Based Image Clustering* (DDC), un modelo que permite tratar inconvenientes de los algoritmos de *Deep Clustering* sobre escoger por anticipado el número de clústeres y resultados de agrupación inestables.

La primera etapa es entrenar un *Deep Convolutional Autoencoder* (CAE) para extraer las representaciones de los descriptores de baja dimensión de los datos de imágenes de alta dimensión, y luego aplica una *T-distributed Stochastic Neighbor Embedding* (t-SNE) para reducir aún más a un espacio de dos dimensiones favoreciendo a los algoritmos de agrupación basados en densidad. En la segunda etapa, se propone una nueva técnica de agrupación basada en densidad para los datos de dos dimensiones para reconocer automáticamente un número apropiado de clústeres con formas arbitrarias. Concretamente, se generan múltiples clústeres locales para capturar las estructuras locales de los clústeres, y luego se combinan por su relación de densidad para formar el resultado de agrupación final.

Los experimentos demuestran que el DDC propuesto logra un desempeño comparable o incluso superior a los métodos de *Deep Clustering* modernos, incluso si el número de clústeres no está dado.

## V. CONCLUSIONES

El objetivo de este trabajo fue realizar una revisión de literatura del *Deep Clustering* para tener un mayor entendimiento sobre su mejor desempeño en la identificación de descriptores en datos de alta dimensión como las imágenes, en comparación con los algoritmos convencionales de agrupación, y conocer las técnicas de agrupación profunda que están a la vanguardia en la clasificación en imágenes. Este estudio se realizó utilizando la metodología SLR, para llevar a cabo una revisión de literatura de forma rigurosa.

Se buscaron artículos científicos digitales publicados en inglés a partir del 2015 en versión completa en el servicio en

línea de información científica *Web of Science* para identificar la literatura importante para cada pregunta de investigación y se realizó la búsqueda de esa literatura en la base de datos bibliográfica *ScienceDirect*; luego se obtuvieron los artículos relevantes y se procedió a realizar una revisión de resúmenes para descartar los elementos que no coincidieran con los criterios de calidad.

Con la revisión de literatura se puede evidenciar como en diversas técnicas y al momento de evaluar el resultado de las agrupaciones logradas, la selección de características que se obtiene mediante las redes neuronales es un elemento esencial para tal éxito.

Con respecto a las técnicas de agrupación profunda, se encontró que están surgiendo métodos nuevos de *Deep Clustering* que están superando a las técnicas *state-of-the-art* con un mejor desempeño en agrupación en imágenes. Esto es interesante teniendo en cuenta la importancia de las aplicaciones que tiene la clasificación de imágenes, por ejemplo, la identificación de cáncer de pulmón en tomografías computarizadas y melanomas en dermatoscopías.

Una de las mayores ventajas del *Deep clustering* es el potencial que tiene para realizar las agrupaciones de forma no supervisada, descubrir patrones significativos a partir de grandes conjuntos de datos sin la presencia de etiquetas es de gran utilidad para muchas aplicaciones en diversos procesos de la ciencia y la industria.

#### REFERENCIAS

- [1] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui y J. Long, «A Survey of Clustering With Deep Learning: From the Perspective of Network Architecture,» *IEEE Acces*, 2018.
- [2] K. Balaji, K. Lavanya y A. Geetha, «Clustering of mixed datasets using deep learning algorithm,» *Chemometrics and Intelligent Laboratory Systems*, 2020.
- [3] A. Abdella y I. Uysal, «A Statistical Comparative Study on Image Reconstruction and Clustering with Novel VAR Cost Function,» *IEEE Access*, vol. 8, pp. 25626-25637, 2020.
- [4] T. Kulkarni, W. Whitney, P. Kohli y J. Tenenbaum, «Deep convolutional inverse graphics network,» *Adv. Neural Inf. Process. Syst.*, pp. 2539-2547, 2015.
- [5] L. Zhou y Y. Wei, «DIC: Deep Image Clustering for Unsupervised Image Segmentation,» *IEEE Acces*, 2020.
- [6] A. dos Santos, D. Matte, G. Gonçalves, H. Pistori y M. Theophilo, «Unsupervised deep learning and semi-automatic data labeling in weed discrimination,» *Computers and Electronics in Agriculture*, 2019.
- [7] M. Al-masni, M. Al-antari, M.-T. Choi y S.-M. Han, «Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks,» *Computer Methods and Programs in Biomedicine*, 2018.
- [8] Z. Zhao, J. Zhao, K. Song, A. Hussain, Q. Du, Y. Dong, J. Liu y X. Yang, «Joint DBN and Fuzzy C-Means unsupervised deep clustering for lung cancer patient stratification,» *Engineering Applications of Artificial Intelligence*, 2020.
- [9] Y. Ren, N. Wang, M. Li y Z. Xu, «Deep density-based image clustering,» *Knowledge-Based Systems*, 2020.