

ANÁLISIS COMPARATIVO DE METODOS DE SEGMENTACION SEMÁNTICA DE NUBES DE PUNTOS CON REDES NEURONALES

Alejandro Espacio, Santiago Salamanca, Pilar Merchán, Emiliano Pérez, Samuel Punzón
Escuela de Ingenierías Industriales, Universidad de Extremadura, 06006 Badajoz.
{alespaci, ssalamanca, pmerchan, emilianoph, spunzon} @unex.es

Resumen

Desde sus inicios, el ámbito de la visión por computador se ha esforzado por recabar de imágenes 2D y espacios 3D la información y el conocimiento que el ser humano puede extraer fácilmente con un solo vistazo. Los métodos más efectivos han apostado por llevar el concepto de inteligencia artificial un paso más allá, optando por unir la extracción de información con el uso de redes neuronales, dando como resultado lo que hoy llamamos aprendizaje profundo, una de las ramas de la computación que más desafíos ofrece a día de hoy. Este artículo no busca sino realizar una recopilación de aquellos métodos que siguen siendo la base, como PointNet, de muchos estudios recientes, así como aquellos que bien podrían suponer nuevas líneas de investigación en el campo de la segmentación semántica de escenas y modelos tridimensionales, dada la importancia y el reto derivado de trabajar con nubes de puntos. Con este objetivo, se realiza una comparativa entre los métodos para quienes deseen introducirse en este campo, considerando aspectos que se creen convenientes para trabajar con nubes de puntos reales.

Palabras clave: Segmentación semántica, estado del arte, redes neuronales profundas, aprendizaje máquina.

1 INTRODUCCIÓN

Actualmente, una de las ramas de la visión por computador que más interés está atrayendo es la de la segmentación semántica. En contraposición a la segmentación tradicional, en la que se detectan en una imagen zonas congruentes entre sí que, posteriormente, son reconocidas mediante algoritmos clásicos de aprendizaje máquina (vecino más próximo, árboles de decisión, SVM...), en la segmentación semántica se realiza el etiquetado de todos y cada uno de los píxeles de la imagen o, como en los casos que se tratan en este trabajo, los puntos de una nube de puntos 3D.

A diferencia de las técnicas clásicas de aprendizaje máquina, que requieren la definición de modelos o características de representación de manera previa al

proceso de aprendizaje, en las redes neuronales profundas (RNP) se aborda la resolución del problema en cuestión sin la necesidad de información contextual inicial. Además, las técnicas de aprendizaje máquina cuentan con varias desventajas con respecto a las RNP: trabajan peor con conjuntos grandes de datos, necesitan el apoyo de la definición de las características y ofrecen, en problemas complejos como el que se trata en este trabajo, resultados más pobres. Es por ello por lo que los mayores avances en segmentación semántica se estén consiguiendo mediante el uso de RNP.

En este artículo se revisan y analizan algunos de los métodos de segmentación semántica de nubes de puntos 3D mediante RNP propuestos hasta el momento. Para ello, en la sección 2 se establece una clasificación de estos métodos atendiendo a si realizan o no tratamiento, o estructuración, de la nube de puntos previo a la red de aprendizaje. Las secciones 3 y 4 agrupan los diferentes métodos siguiendo la clasificación anterior, ordenados cronológicamente para una visión general de la evolución en este campo, y presenta sus funciones, ventajas e inconvenientes. La sección 5 señala las diferencias entre los métodos comentados con el objetivo de que pudiera servir de ayuda a quienes quieran introducirse en el ámbito de la segmentación semántica de nubes de puntos con técnicas basadas en RNP. Cabe destacar la dificultad que supone realizar una comparación entre métodos debido a la disparidad entre las pruebas realizadas en los distintos trabajos. Por último, la sección 6 recoge las conclusiones a las que se ha llegado tras la realización de esta revisión del estado del arte.

2 CLASIFICACIÓN DE LOS MÉTODOS ANALIZADOS

Considerando las técnicas basadas en RNP, distintos artículos [4,20,22] coinciden en clasificar los diferentes métodos en función de los datos que se introducen a las redes neuronales.

Las nubes de puntos, por su propia naturaleza, presentan diversas limitaciones. Una de las más importantes es la distribución irregular de los puntos 3D, lo que obliga a aplicar, en muchos casos, tratamientos de regularización espacial. Otro problema es la falta de estructura topológica de la nube

de puntos. Como primera aproximación para solventar los problemas mencionados se estructuraron y

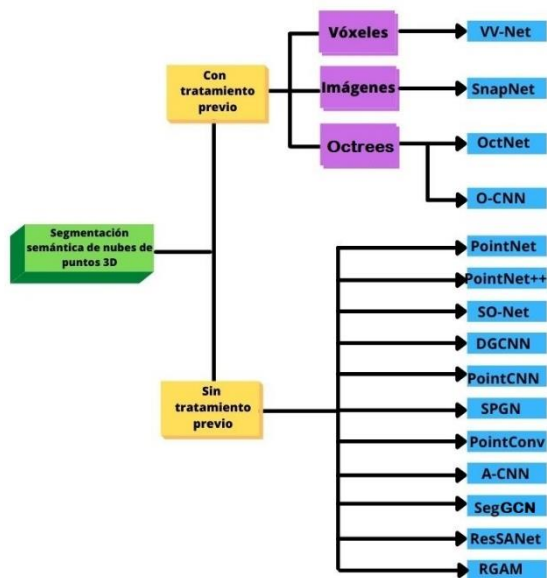


Figura 1: Clasificación de los métodos abordados.

regularizaron las nubes de puntos con vóxeles y octrees. Aunque el uso de estas estructuras de datos soluciona estos problemas, se generan otros, en la mayoría de los casos relacionados con la capacidad de almacenamiento de la información y el coste computacional de los algoritmos.

Por ello, el siguiente paso consistió en generar imágenes a partir de la nube de puntos desde diferentes puntos de vista, e introducir estas imágenes en la red neuronal, normalmente redes neuronales convolucionales (CNN) empleadas en segmentación semántica de imágenes. Los resultados obtenidos utilizando este enfoque eran muy dependientes de la extensión y complejidad de la escena.

Finalmente, la propuesta más reciente consiste en introducir directamente la nube de puntos en la RNP, siendo PointNet [11] la propuesta pionera en este ámbito. Desde entonces, muchos trabajos han mejorado esta técnica, produciendo consecuentemente más y mejores resultados, pero con un amplio rango de mejora, en bancos de pruebas exigentes como S3DIS, ScanNet y ShapeNetPart, siendo estos los más comúnmente empleados por los métodos aquí citados. Como ya se ha comentado, existen dos grandes grupos en los que se podrían separar los métodos de segmentación semántica de nubes de puntos mediante RNP, que son aquellos que necesitan que las nubes sean previamente tratadas o estructuradas y los que no. En la Figura 1 se muestra un esquema de la clasificación de las técnicas aquí recogidas. En amarillo se diferencian los dos grupos mencionados, con y sin tratamiento de las nubes de puntos; en

morado, la clasificación en función del tipo de dato de entrada a la red de aprendizaje; por último, en azul, los diferentes métodos.

A continuación, se explica cada uno de ellos, siguiendo la clasificación propuesta.

3 MÉTODOS CON TRATAMIENTO PREVIO DE LAS NUBES DE PUNTOS 3D

3.1 MÉTODOS BASADOS EN VÓXELES

Estos métodos hacen uso de vóxeles para estructurar las nubes de puntos. Es esta nueva organización de los puntos la que posteriormente se introduce en la red neuronal para ser segmentada. Dividir las nubes en vóxeles conlleva diversas ventajas, como una solución intuitiva al problema de la distribución de los puntos, pero puede suponer otros tantos inconvenientes. Trabajando con vóxeles es posible que algunos queden vacíos o

contengan información no relevante, lo que aumentará el coste de memoria y la carga computacional.

3.1.1 VV-Net

Ante las limitaciones del uso de vóxeles para ordenar los puntos en una nube, los autores de VV-Net propusieron un método a través del cual la información contenida en cada división espacial es mayor que en el uso tradicional de vóxeles. La nube de puntos se reestructura en una distribución ordenada de vóxeles, los cuales son, a su vez, divididos en subvóxeles. Estas últimas divisiones contienen varios puntos que son interpolados, consiguiendo así una representación espacial más suave que con los vóxeles tradicionales. Para compactar de manera efectiva la información anterior, VV-Net entrena un VAE (variational auto-encoder) [5] con el que consiguen mapear la distribución espacial de los puntos presentes en cada subvoxel en un espacio latente.

El método propuesto por Hsien-Yu Meng *et al.* [10] demuestra ser efectivo y robusto en tareas de segmentación semántica con pruebas experimentales en el banco de pruebas de S3DIS, que mejora ligeramente la funcionalidad de los vóxeles, aunque produjo errores puntuales en casos concretos.

Otro ejemplo del uso de vóxeles en este ámbito es SegCloud [14].

3.2 MÉTODOS BASADOS EN OCTREES

A diferencia de los vóxeles, el uso de octrees permite abordar el problema de las divisiones con información relevante. De esta forma, es posible realizar una organización de la nube de puntos a través de octantes

más eficaz y eficiente. El mayor problema de estas técnicas es la posible pérdida de información al dividir la nube de puntos, dado que el resultado sería, al igual que en el uso de vóxeles, una aproximación.

3.2.1 OctNet

Este método, desarrollado por Gernot Riegler *et al.* [13], apuesta por el uso de octrees. Otras propuestas que emplean este tipo de divisiones de escenas tridimensionales hacen uso de un solo octree, mientras que OctNet divide las nubes de puntos en varios octrees de diferentes profundidades y extensión. Así, las zonas con mayor densidad de puntos reciben mayor atención, gracias, además, al uso de redes convolucionales para la extracción de información. A pesar de que este método puede no ser mejor en términos de uso de almacenamiento en comparación con el uso tradicional de octrees, sí mejora la compresión de la información para cada nodo de la estructura. Además, este tipo de estructuras puede ser fácilmente codificado en cadenas de bits, reduciendo así el tiempo de acceso a las celdas de las mismas. OctNet, por ello, permite trabajar con vóxeles de alta resolución, concretamente, resoluciones superiores a 64^3 . Los resultados obtenidos al trabajar con resoluciones altas son mejores que los obtenidos con resoluciones bajas, siendo los mejores resultados los obtenidos con 64^3 .

3.2.2 O-CNN

O-CNN (Octree CNN) [15] propone una idea similar a OctNet: divide la nube de puntos en varias ramificaciones de un solo octree, pero presta toda su atención únicamente en los límites del candidato a objeto 3D, mientras que OctNet trabaja con el interior del mismo, lo que conlleva más carga computacional. O-CNN extrae las características de los octantes de mayor profundidad, que son los que coinciden con los que contienen los límites de la figura con mayor precisión.

Una de las mayores ventajas que proporciona O-CNN es la de procesar diferentes etapas de convolución de manera paralela, permitiendo así su ejecución en la GPU.

O-CNN presenta resultados ligeramente mejores que los obtenidos mediante PointNet, método que se abordará a continuación, en tareas de segmentación parcial de objetos. De igual forma, se puede observar un mejor rendimiento a medida que se aumenta la resolución de los vóxeles, con el máximo en 64^3 .

3.3 MÉTODOS BASADOS EN IMÁGENES

Estos métodos extraen varias imágenes o capturas desde diferentes puntos de vista de la nube de puntos que son posteriormente tratadas en la red neuronal, obteniendo de ellas las características deseadas. Una

vez extraídas las características, se recupera el modelo de la nube de puntos, ahora segmentado.

Por la propia naturaleza de esta técnica, no se trata de una opción óptima en casos donde la escena capturada sea extensa o de gran complejidad, dado que sería necesario realizar una gran cantidad de imágenes para cubrir todas las zonas. Otro problema aparece en las zonas con baja densidad de puntos, donde la información podría ser insuficiente para generar una imagen.

3.3.1 SnapNet

SnapNet [1] es el primer método encontrado en una extensa base bibliográfica que se atrevió con la tarea de segmentación semántica empleando proyecciones de imágenes 2D a partir de nubes de puntos. Este método se desglosa en cuatro fases diferenciadas: preprocesamiento de la nube de puntos y obtención de atributos, generación de proyecciones, segmentación de las imágenes 2D y, por último, vuelta al modelo 3D, produciendo la segmentación tridimensional.

SnapNet, no obstante, sufre diversos problemas. En casos donde las nubes de puntos no son lo suficientemente densas y homogéneas, los resultados pierden interés. Lo mismo ocurre si el conjunto de datos empleado para el aprendizaje de la red neuronal no cuenta con modelos variados y ricos en información. Por tanto, SnapNet puede ser una vía fiable para escenas controladas de poca extensión y una reducida cantidad de elementos.

Desde SnapNet, pocos métodos han surgido con una clara base en el trabajo de imágenes. Un ejemplo es View-GCN [18].

4 MÉTODOS SIN TRATAMIENTO PREVIO DE LAS NUBES DE PUNTOS 3D

Tras los intentos de estructuración sobre las nubes de puntos, PointNet llegó para proponer una nueva manera de trabajar, utilizando directamente las nubes, sin tratamientos previos, salvo las transformaciones necesarias para alcanzar la invariabilidad a permutaciones, como se verá a continuación.

4.1 POINTNET

Desarrollado por C. R. Qi *et al.* [11], PointNet ha sido, desde su origen, la base de muchos de los métodos de visión 3D con nubes de puntos basados en RNP de la actualidad.

PointNet trabaja directamente con los puntos de la nube, a los cuales se les aplican diferentes transformaciones para conseguir esa invariabilidad mencionada anteriormente. A diferencia de los métodos más actuales, PointNet no emplea CNN para

la extracción de las características de los puntos, sino varias etapas de MLP (perceptrón multicapa). Una vez se ha obtenido la característica individual de cada punto, se consigue la característica global de la escena, de nuevo, con una capa MLP, y se concatena con las características individuales previamente extraídas. De este nuevo conjunto, producto de la agregación de características, deriva finalmente otro grupo de características para cada punto.

No son pocos los métodos que han empleado PointNet como raíz principal, los cuales han mejorado considerablemente los resultados en los ámbitos de segmentación, clasificación y detección. Esto es debido a la vanguardista propuesta que supuso en su día, capaz de trabajar con nubes de un millón de puntos, aproximadamente. Sin embargo, el principal problema del que adolece PointNet es la incapacidad de extraer la información topológica de la nube. Diversas pruebas han demostrado que estudiar la relación entre puntos cercanos ofrece mejores resultados en cuanto a la extracción de características y entendimiento de la escena.

4.2 POINTNET++

Un año más tarde, C. R. Qi *et al.* [12] mejoraron la primera versión de PointNet dando lugar a PointNet++. En esta ocasión, la mejorada propuesta hacía uso de una estructura neuronal jerárquica para una mejor extracción de características teniendo en cuenta la información topológica de la nube. Estas se extraen de agrupaciones de puntos cercanos que, posteriormente, generan grupos más grandes con menos elementos. Estos grupos se procesan para conseguir características de mayor nivel y se repite el proceso las veces necesarias para obtener las características de más alto nivel, conformando el grupo de aprendizaje de características.

PointNet++ produce así mejores resultados con respecto a su versión anterior, PointNet. Sin embargo, este método, al trabajar con agrupaciones de puntos cercanos, puede llegar a generar errores en casos donde la densidad de la zona no alcanza ciertos números. Como resultado se pueden observar fallos entre objetos de la escena con una diferencia de escala significativa.

4.3 SO-NET

SO-Net [8] ofrece otro método manteniendo la invariabilidad a permutaciones mediante la realización de un modelo para mapear la distribución espacial de los puntos mediante el uso de un mapa autoorganizado (Self-Organizing Map, SOM), cuyo objetivo es representar la nube en cuestión con un menor número de dimensiones. Posteriormente, la extracción de características se realiza mediante MLP de forma jerárquica tanto a los puntos individuales de la nube

como a los nodos producidos tras aplicar el SOM utilizando el método de los k puntos más cercanos.

Los resultados proporcionados en pruebas para segmentación parcial y clasificación de objetos muestran claras mejoras en comparación con métodos que emplean otras estructuras para la organización de la nube, como O-CNN u OctNet. La ventaja más significativa es el tiempo de entrenamiento, que se encuentra en torno a las 3 horas para modelos sin especificar de ModelNet40, en contraposición a las 20 horas de PointNet++, aunque no se compara con los tiempos que pudieran necesitar los métodos basados en octrees antes mencionados.

4.4 POINTCNN

Para mejorar la pérdida de información relacionada con las nubes de puntos, PointCNN introduce el uso de la transformación X a modo de aprendizaje directamente desde los datos de entrada. De esta forma, el método propuesto por Yangyan Li *et al.* [9] es capaz de obtener la influencia de cada característica por cada punto, así como permutar puntos en un orden canónico. En términos generales, PointCNN es una generalización de las estructuras tradicionales de CNN en el ámbito del aprendizaje de características.

Por tanto, PointCNN consigue mejorar los problemas de densidad no uniforme y desorden de las nubes de puntos. A pesar de ello, el uso de la transformación X necesita un conocimiento exhaustivo del operador cuando se pretende aplicar a una red neuronal. Por otro lado, presenta resultados prometedores en segmentación en comparación con métodos como PointNet/PointNet++, SO-Net, DGCNN [3] o SGPN [16]. En clasificación, PointCNN ofrece peores resultados con respecto a otros métodos.

4.5 SGPN

Se basa en el uso de las estructuras PointNet/PointNet++ aplicadas a cada uno de los puntos de la nube de entrada de forma que extrae tanto características locales como globales. Posteriormente, SGPN (Similarity Group Proposal Network) se divide en tres ramas claramente diferenciadas: generación de una matriz de similitud, generación de un mapa de confianza y segmentación semántica. Partiendo de la idea de que puntos pertenecientes a un mismo objeto son más propensos a contar con un vector de características similar y a estar situados cerca los unos de los otros, esta matriz de similitud genera diferentes candidatos para ser considerados objetos. A estos candidatos se les aplica el mapa de confianza con el objetivo de escoger el que mejor resultados aporte, siendo el elegido el que se somete a continuación a la red de segmentación semántica.

A pesar de la flexibilidad y eficacia de SGPN, dado que la matriz de similitud aumenta de manera

cuadrática conforme aumenta el tamaño de la nube de puntos procesada, el método ideado por W. Wang *et al.* no es adecuado para trabajar escenas de más de 100.000 puntos siendo esta una clara limitación para uso con datos reales adquiridos con escáneres de corto o largo alcance. Los resultados de las pruebas aportados por los responsables, aunque positivos, no son mejores que los de los métodos actuales, aunque su simpleza es una clara ventaja.

4.6 DGCNN

Yue Wang *et al.* [3] introducen dos conceptos interesantes para las redes neuronales de grafos (Graph Neural Networks, GNN). Por un lado, aplican una novedosa operación denominada EdgeConv, con el que se consigue obtener las características locales de los puntos, y la relación de estos con sus puntos más cercanos, mientras se respeta la invariancia a permutaciones. La otra introducción es DGCNN (Dynamic Graph CNN), método que, a diferencia de las redes neuronales de grafos más empleadas, actualiza el resultado del grafo tras cada capa de la red, es decir, el resultado no es fijo hasta que no se termina el proceso.

DGCNN presenta grandes resultados en términos de OA (Overall Accuracy) frente a otros métodos como PointCNN. Sin embargo, aunque innovadora, esta propuesta cuenta con aspectos mejorables. LDGCNN (Linked DGCNN) [23] es un ejemplo que mejora la anterior estructura.

4.7 POINTCONV

PointConv [19] parte de la idea de considerar la operación de convolución como una aproximación discreta aplicando una variación del método de Monte Carlo. El peso de las características se obtiene a través de perceptrones multicapa y funciones de densidad, aunque dichos valores son recalculados posteriormente con PointConv, consiguiendo así una mejora en resultados y tiempo de computación en comparación con PointNet en tareas de segmentación semántica y PointCNN en clasificación.

4.8 A-CNN

El método propuesto por Artem Komarichev *et al.* [6] introduce una nueva variante de operador convolucional. En este caso, A-CNN (Annularly Convolutional Neural Network) propone el uso de estructuras anulares para la captación de puntos, y la consiguiente formación de agrupaciones, a diferentes escalas. Estas estructuras pueden ser bien regulares o dilatadas, siendo la diferencia entre ellas un espacio vacío intermedio entre los distintos niveles. Ambas aproximaciones evitan el solapamiento o duplicación de puntos, mejoran la extracción de características geométricas y permiten trabajar con zonas más

extensas empleando kernels del mismo tamaño sin necesidad de aumentar el número de parámetros para las etapas de convolución.

Como resultado se obtiene un método que produce una ligera mejora en segmentación semántica, segmentación parcial y clasificación de objetos en escenas de mayor extensión que otras técnicas basadas en métodos multiescala, como puedan ser PoinNet/PoinNet++, PointCNN, 3P-RNN [21], SO-Net o SGPN. La novedad de su propuesta puede suponer una interesante línea de investigación para futuros proyectos.

4.9 SEGGCN

Huan Lei *et al.* [7] proponen una serie de coeficientes para el kernel borroso, una solución a los problemas de los kernels esféricos, así como el SegGCN (encoder-decoder graph convolutional network), con el fin de aplicar de manera efectiva el mencionado kernel, siendo capaz de procesar hasta un millón de puntos de la nube por segundo.

La técnica propuesta por SegGCN puede ser una interesante línea de investigación para futuros proyectos, aunque los resultados proporcionados, en comparación con otros métodos como SPG, PointConv o PointCNN, en bancos como S3DIS o ScanNet, son muy similares.

4.10 RESSANET

Los autores de ResSAnet [24] propusieron recientemente una técnica similar a la convolución, con la salvedad de que emplea figuras geométricas primitivas para la extracción de características geométricas de las nubes de puntos respetando la invariabilidad a permutaciones. Este método, además, introduce el uso de dos módulos, Res-SA y Res-SA-2, que ofrecen una nueva vertiente a la hora de realizar saltos de conexiones para la adición de características multiescala y multinivel.

Se puede observar una ligera mejora en los resultados mostrados en comparación con otros métodos como PointNet o DGCNN. Además, ResSAnet presenta una reducción significativa en términos de carga computacional y número de parámetros de la red neuronal. Sin embargo, una desventaja es la necesidad de ajustar ciertos parámetros durante de proceso de entrenamiento, haciéndolo menos independiente que otros propuestos en esta revisión.

4.11 RGAM

Métodos como SegGCN, SO-Net y DNNSP [17] fallan a la hora de tener en cuenta la información contextual de la escena, clave para la obtención de características globales y su posterior segmentación. A raíz de lo anterior surgió RGAM (ring grouping neural network with attention module) [2]. RGAM es

capaz de extraer tanto características locales como globales teniendo en cuenta las vecindades colindantes, mejorando el cálculo de puntos repetidos y reduciendo el tiempo de computación.

En términos generales, mejora los resultados obtenidos de segmentación semántica en bancos de prueba como S3DIS y NYU-V2. No obstante, debido a que en esta red los autores proponen el uso de una mini-PointNet en la etapa inicial de aprendizaje, no es capaz de trabajar con estructuras geométricas complejas. En el citado campo, se trata de la propuesta más reciente y que mejores resultados ofrece para escenas interiores en comparación con varios métodos, como PointNet, PointNet++, SegCloud, SPGN y DGCNN.

5 COMPARATIVA

En esta sección se introduce una breve comparación entre los distintos métodos indicados anteriormente. Entre los aspectos a comparar se encuentran la disponibilidad del código para acceso y testeo de cada método, si se pueden aplicar para espacios interiores, exteriores, no convencionales (NC) u objetos. Entiéndase por escenas no convencionales aquellas que presenten estructuras de geometría compleja, por ejemplo, una bóveda. Además, también se indica si los métodos en cuestión incluyen en sus respectivos artículos alguna mención a tiempos de entrenamiento. La Tabla 1 recoge la comparación entre los diferentes escenarios, mientras que la Tabla 2 reúne los aspectos relacionado con el código y el tiempo mencionados. Es necesario destacar la dificultad a la hora de comparar los diversos métodos, debido, en algunos casos, a la falta de información, y, en otros, a la disparidad en las pruebas realizadas.

Por un lado, se cree indispensable conocer el número de puntos con los que pueden llegar a trabajar estas técnicas, dado que las nubes de puntos pueden llegar a contener millones de estos. Por otro, ciertos métodos, como A-CNN, realizan pruebas en bancos de pruebas como S3DIS o ShapeNet-part mientras que SO-Net trabaja con otros como ModelNet40 o MNIST.

Lo mismo ocurre con las condiciones en las que se realizan las pruebas, con conjuntos de puntos que varían en número y distribución, y que difícilmente se pueden comparar. El hardware empleado en los ensayos de cada método, de igual forma, es un factor determinante en los resultados obtenidos, haciendo muy complicada la comparación de los resultados en términos cuantitativos. Todas estas limitaciones hacen que la comparación entre métodos se convierta en una tarea compleja.

Las tablas 1 y 2, por tanto, pretenden ofrecer una primera aproximación para quienes deseen introducirse en el campo de la segmentación semántica en relación con las nubes de puntos 3D. A

continuación, se exponen las conclusiones a las que se ha llegado.

1. Muchos de estos métodos realizan pruebas con nubes del orden de, como máximo, 5000 puntos, una cantidad ínfima en comparación con la resolución de las nubes de puntos reales. Por tanto, sería interesante realizar pruebas en nubes de puntos reales, para la obtención de resultados más representativos de la realidad.
2. Diversos autores aquí mencionados ponen de manifiesto su intención de mejorar sus respectivos métodos en el ámbito de las escenas exteriores y de gran escala, por lo que bien podría tratarse de una línea de investigación de gran interés, siendo una posible aplicación la reconstrucción y preservación de emplazamientos históricos.
3. La mayoría de pruebas realizadas, las cuales vienen indicadas en los documentos de cada método, muestran grandes resultados en escenas tradicionales, con paredes rectas y elementos típicos de interiores, y segmentación parcial de objetos, sin ninguna mención a las estructuras no convencionales antes definidas. Sin embargo, si hubiera que escoger una opción para futuras investigaciones en este ámbito esa sería A-CNN, dado los buenos resultados obtenidos en términos de estructuras anulares.

Tabla 1. Comparación entre métodos para escenas interiores, exteriores, no convencionales y objetos.

Método	Interior	Exterior	NC	Objeto
VV-Net	Sí	-	-	Sí
SnapNet	Sí	Sí	-	-
OctNet	Sí	Sí	-	Sí
O-CNN	-	-	-	Sí
PointNet	Sí	-	-	Sí
PointNet++	Sí	-	-	Sí
SO-Net	-	-	-	Sí
PointCNN	-	-	-	Sí
SGPN	Sí	-	-	Sí
DGCNN	Sí	-	-	Sí
PointConv	Sí	-	-	Sí
A-CNN	Sí	-	Sí	Sí
SegGCN	Sí	-	-	-
ResSAnet	-	-	-	Sí
RGAM	Sí	-	-	-

Tabla 2. Comparación de métodos para disponibilidad del código, número de puntos máximo o tiempo de entrenamiento.

Método	Código	Tiempo
VV-Net	-	-

SnapNet	Sí	-
OctNet	Sí	-
O-CNN	Sí	-
PointNet	Sí	Sí
PointNet++	Sí	Sí
SO-Net	Sí	Sí
PointCNN	Sí	Sí
SGPN	Sí	Sí
DGCNN	Sí	-
PointConv	Sí	Sí
A-CNN	Sí	-
SegGCN	Sí	-
ResSAnet	-	-
RGAM	-	-

6 CONCLUSIÓN

Con este documento se ha realizado una revisión del estado del arte de los métodos empleados para la segmentación semántica de nubes de puntos basados en el uso de RNP, una de las técnicas más empleadas durante los últimos años y que mejores resultados aporta a la escena de la visión por computador. Asimismo, se han clasificado los distintos métodos por su necesidad de tratar las nubes previamente o no. Dada la disparidad de métodos y las pruebas realizadas en diferentes bancos de pruebas, la comparación directa entre técnicas resulta una tarea compleja. A pesar de los resultados aportados por los diferentes autores, la elección de un método concreto para la segmentación semántica, clasificación y detección de objetos de nubes de puntos 3D no es trivial.

No son pocas las aproximaciones realizadas para llevar a cabo las tareas mencionadas de la mejor manera posible, por lo que se cree que existe un amplio margen de mejora en este campo. Las últimas técnicas ofrecen resultados prometedores y futuras líneas de investigación, como tareas de segmentación semántica en escenas exteriores o no convencionales, pueden ser determinantes para conseguir dominar la segmentación semántica de nubes de puntos 3D.

Agradecimientos

Este trabajo ha sido financiado por el proyecto titulado “Procesamiento de datos masivos de escáneres 3D: Segmentación, reconocimiento y modelado geométrico de edificaciones no convencionales”, con referencia PID2019-108271RB-C32. AEI. 10.13039/501100011033; y el proyecto “Tecnologías 3D para la resiliencia, la recuperación y la sostenibilidad del Patrimonio Cultural Extremeño”, con referencia IB20172, de la Junta de Extremadura y el Fondo Europeo de Desarrollo Regional “Una manera de hacer Europa”.

English summary

COMPARATIVE ANALYSIS OF POINT CLOUD SEMANTIC SEGMENTATION METHODS BASED ON NEURAL NETWORKS

Abstract

During the recent years, computer vision has put efforts into adding information to 2D images and 3D scenes the same way humans extract knowledge just by looking at them. Regarding this, many methods have arisen in the past decade, some of which are machine learning-based. The more efficient ones, however, have been aiming to further improve the concept of artificial intelligence by fusing information extraction with neural networks, what it is call as deep learning techniques, a recent and challenging field in computer vision. This paper tries to survey the methods that still are the basis of the newest researches, such as PointNet, as well as those techniques that could reach new state-of-the-art results in scene and 3D models semantic segmentation based on point clouds. A comparative analysis between methods is performed for this purpose, taking into account some characteristics considered important.

Keywords: semantic segmentation, machine learning, deep learning, state-of-the-art.

Referencias

- [1] Boulch, A., Guerry, J., Le Saux, B., Audebert, N., (2018) SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks, *Comput. Graph.* 71 189–198.
- [2] Chen, X.-T., Li, Y., Fan, J.-H., Wang, R., (2021) RGAM: A novel network architecture for 3D point cloud semantic segmentation in indoor scenes, *Inf. Sci. (Ny)*. 571 87–103.
- [3] Cui, Y., Liu, X., Liu, H., Zhang, J., Zare, A., Fan, B., (2021) Geometric attentional dynamic graph convolutional neural networks for point cloud analysis, *Neurocomputing*. 432 300–310.
- [4] Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M., (2020) Deep Learning for 3D Point Clouds: A Survey, *IEEE Trans. Pattern Anal. Mach. Intell.* 8828 1–1.
- [5] Kingma, D.P., Welling, M., (2014) Auto-encoding variational bayes, *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc.* 1–14.
- [6] Komarichev, A., Zhong, Z., Hua, J., (2019)

- A-CNN: Annularly convolutional neural networks on point clouds, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2019-June 7413–7422.
- [7] Lei, H., Akhtar, N., Mian, A., IEEE Computer Society, 2020SEG-GCN: Efficient 3D point cloud segmentation with fuzzy spherical kernel, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, : pp. 11608–11617.
- [8] Li, J., Chen, B.M., Lee, G.H., IEEE Computer Society, 2018SO-Net: Self-Organizing Network for Point Cloud Analysis, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, : pp. 9397–9406.
- [9] Li, Y., Bu, R., Di, X., (2018) PointCNN : Convolution On X -Transformed Points,.
- [10] Meng, H.Y., Gao, L., Lai, Y.K., Manocha, D., Institute of Electrical and Electronics Engineers Inc., 2019VV-net: Voxel VAE net with group convolutions for point cloud segmentation, in: *Proc. IEEE Int. Conf. Comput. Vis.*, : pp. 8499–8507.
- [11] Qi, C.R., Su, H., Mo, K., Guibas, L.J., (2016) PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation,.
- [12] Qi, C.R., Yi, L., Su, H., Guibas, L.J., (2017) PointNet++: Deep hierarchical feature learning on point sets in a metric space, *Adv. Neural Inf. Process. Syst.* 2017-Decem 5100–5109.
- [13] Riegler, G., Ulusoy, A.O., Geiger, A., Institute of Electrical and Electronics Engineers Inc., 2017OctNet: Learning deep 3D representations at high resolutions, in: *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, : pp. 6620–6629.
- [14] Tchapmi, L., Choy, C., Armeni, I., Gwak, J., Savarese, S., (2018) SEGCloud: Semantic segmentation of 3D point clouds, *Proc. - 2017 Int. Conf. 3D Vision, 3DV 2017.* 537–547.
- [15] Wang, P.S., Liu, Y., Guo, Y.X., Sun, C.Y., Tong, X., (2017) O-CNN: Octree-based convolutional neural networks for 3D shape analysis, *ACM Trans. Graph.* 36.
- [16] Wang, W., Yu, R., Huang, Q., Neumann, U., (2018) SGPNet: Similarity Group Proposal Network for 3D Point Cloud Instance Segmentation, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2569–2578.
- [17] Wang, Z., Zhang, L., Zhang, L., Li, R., Zheng, Y., Zhu, Z., (2018) A deep neural network with spatial pooling (DNN-SP) for 3-D point cloud classification, *IEEE Trans. Geosci. Remote Sens.* 56 4594–4604.
- [18] Wei, X., Yu, R., Sun, J., (2020) View-GCN: View-based graph convolutional network for 3D shape analysis, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 1847–1856.
- [19] Wu, W., Qi, Z., Fuxin, L., IEEE Computer Society, 2019PointConv: Deep convolutional networks on 3D point clouds, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, : pp. 9613–9622.
- [20] Xie, Y., Tian, J., Zhu, X.X., (2020) Linking Points With Labels in 3D: A Review of Point Cloud Semantic Segmentation, *IEEE Geosci. Remote Sens. Mag.* 8 38–59.
- [21] Ye, X., Li, J., Huang, H., Du, L., Zhang, X., (2018) 3D recurrent neural networks with context fusion for point cloud semantic segmentation, *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics).* 11211 LNCS 415–430.
- [22] Zhang, J., Zhao, X., Chen, Z., Lu, Z., (2019) A Review of Deep Learning-Based Semantic Segmentation for Point Cloud, *IEEE Access.* 7 179118–179133.
- [23] Zhang, K., Hao, M., Wang, J., de Silva, C.W., Fu, C., (2019) Linked Dynamic Graph CNN: Learning on Point Cloud via Linking Hierarchical Features, 1–8.
- [24] Zhu, X., Zhang, Z., Ruan, J., Liu, H., Sun, H., (2021) ResSANet: Learning geometric information for point cloud-processing, *Sensors.* 21.



© 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution CC BY-NC-SA 4.0 license (<https://creativecommons.org/licenses/by-ncsa/4.0/deed.es>).