



Universitat Oberta
de Catalunya

Universitat Oberta de Catalunya

MÁSTER DE CIENCIA DE DATOS

PEC4 - ESTADO DEL ARTE

Deep Learning

Autor:

Mario Ubierna San Mamés

11 de junio de 2022

Índice general

Índice general	1
Resumen	3
1.1. Referencia del artículo	3
1.2. Descripción de la temática	3
1.3. Novedades que presenta el artículo	3
1.4. Resumen de la parte experimental	4
1.5. Conclusiones	4
Bibliografía	5

Resumen

1.1. Referencia del artículo

El artículo seleccionado es " *Deep Learning Based Automatic Video Annotation Tool for Self-Driving Car*", cuyos autores son *N.S.Manikandan, K.Ganesan*. La fecha de publicación del artículo data del *19 de abril del 2019*, y fue publicada por el grupo *TIFAC-CORE in Automotive Infotronics* perteneciente al Instituto Tecnológico de Vellore [1].

1.2. Descripción de la temática

El *Deep Learning* es un concepto muy usado actualmente, ha tenido un crecimiento exponencial en los últimos años. Es tal el impacto, que se hace uso de este campo para la detección de diferentes elementos encontrados durante la conducción de coches autodirigidos.

Esta tarea es fundamental en los tiempos en los que los coches autodirigidos van adquiriendo más importancia con empresas como *Tesla*.

El objetivo de este artículo es mostrar cómo es la identificación de diferentes objetos, la clasificación de los mismos, la detección de los carriles y el seguimiento de la trayectoria de todos los elementos anteriores.

Para ello se hace uso de técnicas de *deep learning* y herramientas de vídeo, con el fin de capturar cada una de las imágenes en tiempo real y clasificarlas.

1.3. Novedades que presenta el artículo

Inicialmente cuando se empezó a desarrollar los coches autodirigidos, la detección de objetos se realizaba de forma manual y para ello se empleaban programas *open source*. Realizar la detección de objetos de forma manual es muy costosa en términos temporales y económicos.

Es aquí donde entra las novedades del *deep learning* y en consecuencia de este artículo, hacer este proceso de anotación en tiempo real más barato y preciso que de forma manual.

Para alcanzar el objetivo que se plantea en el artículo hay dos estrategias diferentes, hacer uso de métodos de anotación semi-automáticos o totalmente automáticos.

Cabe destacar que hay proyectos similares tal y como se indican en el artículo, pero éstos en su mayoría tienen un enfoque de detección semi-automática. En el presente estudio se busca ir un paso más allá, hacer que la tarea de detección sea completamente automática.

Con el fin de lograr este punto, se hace uso de un trabajo previo de *Zhujun Xiao*, el cual diseñó un sistema de generación de imágenes con autodetección.

1.4. Resumen de la parte experimental

En cuanto a la detección de objetos se han usado tanto *YOLO* como *Retinanet*. El modelo ganador ha sido *Retinanet*, consigue una mayor precisión, una mayor exactitud y una menor pérdida.

En lo que respecta a las propiedades de clasificación se comparan los modelos *VGG-19* y *Retinanet*. El modelo seleccionado como mejor es *Retinanet*, presenta una menor pérdida y una mayor precisión media tanto en el entrenamiento como en la validación.

Por otro lado, para la detección de líneas se hizo uso de *Udacity* y *LaneNet*. *Udacity* tuvo dificultades en la detección de líneas de carreteras asiáticas. Debido a esto se hizo uso de *LaneNet*.

Finalmente, para el seguimiento de objetos se realizó el estudio de *Udacity* y *Deep SORT*. Fue este último el que consiguió mejores resultados, por lo que se utilizó este modelo en la versión final.

1.5. Conclusiones

En este artículo podemos ver el cómo diferentes algoritmos de *deep learning* son capaces de realizar diferentes tareas de identificación sobre un vídeo en tiempo real.

A la vista de los resultados obtenidos, se consigue una exactitud media del 83 %, es decir, todos los modelos en su conjunto son capaces de identificar y anotar de forma automática el 83 % de los objetos analizados, si la anotación manual es el 100 % solamente es un 17 % peor.

Aunque hay una pérdida en la exactitud del 17 % respecto a la clasificación manual, el tiempo requerido es muchísimo menor. Para hacernos una idea, el entrenamiento con CPU tarda 43 minutos en entrenarse, con GPU 2 minutos y medio, pero de forma manual se tarda 3060 minutos.

En conclusión, aunque la exactitud obtenida para los modelos de *deep learning* es menor, el tiempo necesario para realizar la misma tarea haciendo uso de redes neuronales artificiales frente a una labor manual es muy diferente, es casi 1200 veces más rápido la solución que se presenta en este artículo que hacer la misma tarea de forma manual.

Por lo tanto, vemos que sí que se produce un gran avance en esta área. El objetivo en líneas futuras podría ser mejorar la exactitud obtenida en este proyecto.

Bibliografía

- [1] N. S. Manikandan and K. Ganesan, “Deep learning based automatic video annotation tool for self-driving car.” type: article.