



Aprendizaje profundo para visión por computadora y reconocimiento de escenas
Helisa Dhamo. Laboratorio «Noah's Ark» de Huawei. Reino Unido.
Curso en inglés.

Breve resumen de la materia

El aprendizaje profundo es un método de aprendizaje automático que se basa en aprender patrones a partir de un conjunto grande de ejemplos. Es un método particularmente interesante para problemas complejos, para los cuales los datos (experiencia) están ampliamente disponibles, pero formular una solución analítica no es factible. En este curso, exploraremos los conceptos básicos de la inteligencia profunda y la visión por computadora. Mostraremos a través de conferencias teóricas y ejemplos prácticos cómo crear y entrenar modelos de inteligencia profunda según la tarea (detección de objetos, segmentación de instancias, predicción de relaciones entre objetos) y la modalidad de datos (imagen, video, 3D). El curso concluirá con una introducción a algunos temas avanzados y una discusión sobre tendencias recientes.

Objetivos del curso

El primer objetivo de este curso es familiarizar a los estudiantes con los conceptos de la inteligencia profunda y la visión por computadora. Además, mostraremos cómo la inteligencia profunda se puede utilizar en diversas aplicaciones de visión por computadora relacionadas con la percepción. De esta manera, los estudiantes podrán comprender qué algoritmos son más adecuados en función de la representación de datos, por ejemplo, imágenes frente a videos frente a 3D. Además, aprenderán sobre una variedad de tareas de visión por computadora, incluyendo la detección de objetos y la segmentación de instancias/semántica. El objetivo es que al final del curso, los estudiantes puedan determinar qué arquitecturas de redes son adecuadas para cada tarea y por qué. A través de una combinación de conferencias teóricas y ejemplos prácticos de programación, adquirirán experiencia práctica en estos temas. Una vez que adquieran los fundamentos relevantes en estos temas, el objetivo es fomentar su pensamiento crítico sobre las tendencias actuales y los problemas abiertos importantes para el futuro. Con este propósito, hacia el final del curso se considerarán temas más avanzados como discusiones abiertas entre los estudiantes, como la comprensión de escenas conscientes del contexto a través de gráficos de escenas, generación de imágenes, etc. El desempeño de los estudiantes se evaluará a través de un conjunto de ejercicios prácticos.

Programa

- Motivación: demostración de aplicaciones de la inteligencia profunda en visión por computadora.

- Introducción a los fundamentos de la visión por computadora.
- Codificación básica utilizando la biblioteca OpenCV.
- Fundamentos de la inteligencia profunda y redes neuronales convolucionales.
- Introducción a un marco común de inteligencia profunda (PyTorch). Codificación básica en PyTorch.
- Inteligencia Profunda para la visión por computadora en 2D y percepción.
- Representaciones en 3D en la visión por computadora. Inteligencia profunda para la visión en 3D.
- Tema avanzado: comprender escenas dependientes del contexto utilizando gráficos de escenas.
- Resumen de tendencias muy recientes en la visión por computadora.
- Ideas para futuras investigaciones, consideraciones éticas y aplicaciones importantes.

Prerrequisitos

Nociones básicas de álgebra lineal y análisis (nivel CBC), probabilidad y estadística y programación (preferiblemente en Python).

Opcional: nociones de Machine Learning, procesamiento de señales, grafos.

Bibliografía

- Goodfellow, I. J., Bengio, Y., and Courville, A. (2016). Deep Learning. MIT Press, Cambridge, MA, USA. <http://www.deeplearningbook.org>.
- Hartley, R. I. and Zisserman, A. (2004). Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0521540518, second edition.
- He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision (ICCV).
- Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Wald, J., Dhano, H., Navab, N., and Tombari, F. (2020). Learning 3d semantic scene graphs from 3d indoor reconstructions. In Conference on Computer Vision and Pattern Recognition (CVPR).
- Xu, D., Zhu, Y., Choy, C., and Fei-Fei, L. (2017). Scene graph generation by iterative message passing. In Computer Vision and Pattern Recognition (CVPR).