idw - Informationsdienst Wissenschaft
Nachrichten, Termine, Experten

Pressemitteilung

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, DFKI
Jeremy Gob

25.03.2024
http://idw-online.de/de/news830828

Forschungsergebnisse, Wissenschaftliche Tagungen
Gesellschaft, Informationstechnik, Mathematik, Sprache / Literatur
überregional

Viewing the world through human eyes: DFKI-Technology shows robots how it's done

How can we teach organic visual orientation to machines? That is exactly what scientists at the German Research Centre of Artificial Intelligence (DFKI) are trying to figure out. Some of their latest results will be presented at the annual Conference on Computer Vision and Pattern Recognition (CVPR) in Seattle, USA, by DFKI-researchers of the Department Augmented Vision. Their final objective: Breaching the gap between human- and machine-perception. DFK-scientist try to resolve this issue through a Multi-Key-Anchor Scene-Aware Transformer for 3D Visual Grounding (MiKASA).

We, as, humans intuitively learn the concepts of meaning and the suptle elements of semantics. This way we are able to deduce a multitude of intentions and references and match them to real objects in our environment - no matter the actual verbal expression. Machines do not posses this ability yet or at least can not replicate our intuitive semantic repertoire. DFK-scientist try to resolve this issue through a Multi-Key-Anchor Scene-Aware Transformer for 3D Visual Grounding (MiKASA). MiKASA enables the understanding of highly complex spacial relations and object attributes and further allows machines to identify said objects and recognize them through semantics.

Context is key

„For instance, when we come across a cuboid-shaped object in a kitchen, we may naturally assume that it is a dishwasher. Conversely, if we spot the same shape in a bathroom, it is more plausible that it is a washing machine", so Alain Pagani of the DFKI explains. Meaning is contextual and this information seems to be crucial in determining the true identity of an object and adds a nuanced understanding of our surroundings. Thanks to MiKASAS scenenaware object-encoder, machines are able to interpret information through the imminent surroundings of any reference-object and can identify and define said object accurately.

Unfortunately, machines and programs face another challenge in understanding spacial relations like we do: perspective! "The chair in front of the blue-lit monitor" can transform into the "chair behind the monitor", depending on ones point of view. The chair remains the same, yet its position and orientation in the environment may change. In order for machines to understand, that both described chairs are in fact the same object, MiKASA incorporates a so called "multi-key-anchor concept". It transmits the specific coordinates of anchor points in the field of view in relation to the targeted object and evaluates the importance of nearby object through text-descriptions. Through this method, semantic references aid in localization. The logic behind: a chair is typically faced towards a table or is positioned at a wall. The presence of a table and wall therefore define how a given chair would be orientated in the room indirectly.

The merger of languagemodells, acquired knowledge of semantics and object-recognition in 3D space, leads MiKASA to an accuracy of 78,6 percent (Sr3D Challenge), being a ten percent hit rate increase in the field of object-recognition in comparison to current other technologies!

„Seeing“ does not equal „understanding“

The baseline for eventual understanding is perception. This is why thousands of sensors are needed to gather data and merge them into a holstici representation of the gathered information. Without this cluster of data, spacial orientation becomes impossible - for humans and robots alike. Thankfully, the human brain is able to gather visual information through both our eyes and combine them into one consistent picture, omitting overlaps and duplicate sensoriy information automatically. Replicating this ability for machines proves to be difficult, but researchers at the DFKI may have found an efficient solution: the Partial Graph Matching Network with Semantic Geometric Fusion for 3D Scene Graph Alignment and its Downstream Tasks (SG-PGM).

The alignment between so called 3D scene graphs offers the foundation for a multitude of use cases. It supports point cloud registration - and helps robots to navigate the world. In order to enable the same in dynamic scenes with unpredictable noize, SG-PGM links those visualisations with a neural network. "Our program recycles geometric elements which have been learned through point cloud registration and associates the clustered geometric data points with semantic attributes at node level", so researcher Alain Pagani, Department Augmented Vision at the DFKI. Essentially, a group of points is given a semantic meaning (for example the meaning: "blue chair in front of a monitor"). The same group can then be recognized in another graph and does not have to be visualized again in the final representation of the scene.

Through this method SG-PGM accomplishes to detect overlaps with high precision and generates as exact of an representation as possible. This means: Better orientation for robots in their specific environment and accurate localization of objects. In order to honor this achievement, the organizers of the annual CVPR in Seattle have granted a placement to the corresponding paper.

Didier Stricker, Department-director of Augmented Vision at the DFKI and his team will be present with six different publications, for example, covering technologies for 3D object-recognition through variable text descriptions and holistic scene-encoding.

wissenschaftliche Ansprechpartner:

Didier.Stricker@dfki.de, Prof. Dr. Didier Stricker, AV - Lead
Alain.Pagani@dfki.de, Dr.-Ing. Alain Pagani, AV

URL zur Pressemitteilung: http://From 17th until 21st of June the CVPR 2024 will take place at the Seattle Convention Center and is renowned for being one of the most important events in machine-pattern-regocnition. Out of thousands of submissions the most relevant technological methods and developers have been invited to the conference in Seattle.

A robot, trying to view the scene as we would
DFKI Kaiserslautern (Stockfoto)
DFKI Kaiserslautern