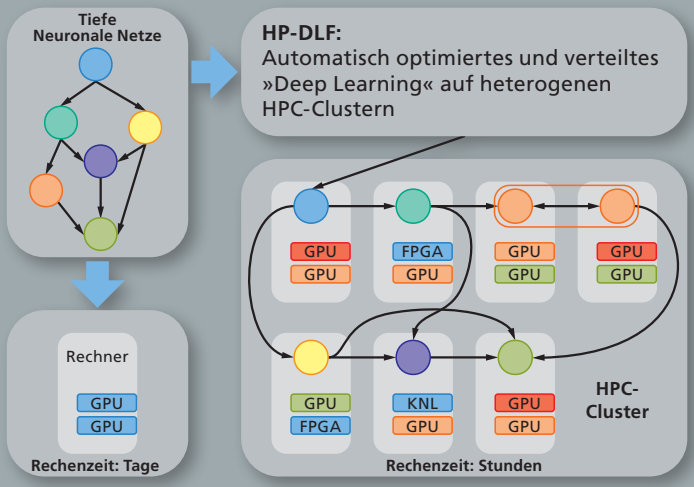

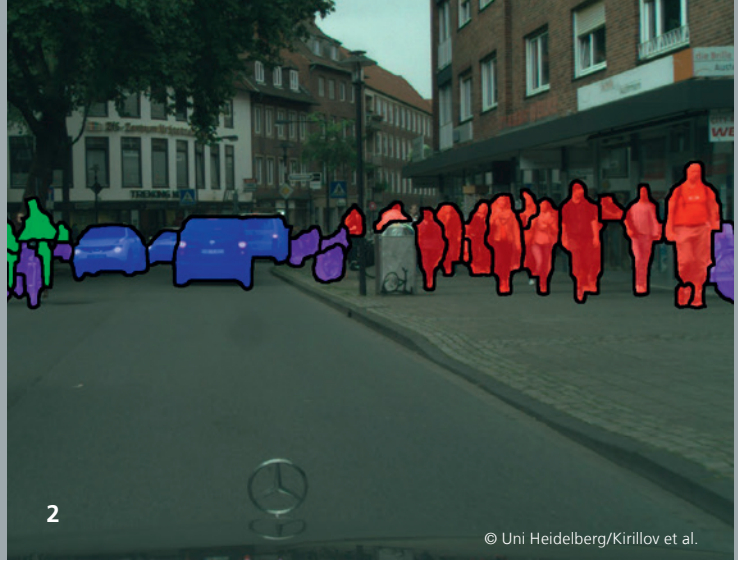

HPC FÜR MASCHINELLES LERNEN: HIGH PERFORMANCE DEEP LEARNING FRAMEWORK

Künstliche neuronale Netze haben sich in den vergangenen Jahren in vielen Bereichen des maschinellen Lernens durchgesetzt. So bilden sie beispielsweise auf dem Gebiet der Computervision, der Sprach- und Texterkennung als auch der maschinellen Übersetzung den Stand der Technik. Grund für ihren Erfolg ist u.a. ihre Fähigkeit, hochgradig komplexe Zusammenhänge zwischen rohen Eingabedaten und den klassifizierenden Ausgabedaten (den Labeln) herstellen zu können.

1 *HPC ermöglicht Deep Learning ohne Speichergrenzen.*

2 *Große Datenmengen spielen beim Autonomen Fahren eine besondere Rolle.*

Dafür benötigen sie häufig mehrere Millionen freie Parameter, die während des Trainings des Netzes verändert, d. h. gelernt werden. Die große Anzahl dieser sogenannten Gewichte führt allerdings dazu, dass das Training eines einzelnen neuronalen Netzes häufig mehrere Tage oder sogar Wochen dauern kann. Deshalb ist es wünschenswert, diese Algorithmen durch den Einsatz von Supercomputern stark skalierbar zu machen. Dies würde im Idealfall bedeuten, dass eine Verdopplung der Anzahl parallel geschalteter Computer zu einer Halbierung der Laufzeit des Algorithmus führen würde.

Kleine neuronale Netze oder wenige Dateien?

Ein weiteres Problem, auf das man mit neuronalen Netzen trifft, ist ihr großer Bedarf an Hauptspeicher. Das hat zur Folge, dass man auf einem einzelnen Rechner nur relativ kleine neuronale Netze trainieren kann, oder sich bei der Menge der zum Lernen verwendeten Daten beschränken muss. Weder das eine noch das andere ist erstrebenswert, weil die Kapazität, d.h. die Lernfähigkeit des Netzes, reduziert wird. Vielmehr ist es wünschenswert, mit der doppelten Anzahl an Rechnern auch Netze von doppelter Größe trainieren zu können. Dies bezeichnet man im Parallelen Rechnen als schwache Skalierbarkeit.

Hohe Skalierbarkeit mit GPI-Space

Sowohl schwache als auch starke Skalierbarkeit beim Training von neuronalen Netzen zu ermöglichen, ist Gegenstand des BMBF-Projekts »High Performance Deep Learning Framework« (HP-DLF). Wir sind dabei vor allem daran interessiert, neuronale Netze beliebiger Größe konstruieren zu können und einen einfachen Zugang zu existierenden und zukünftigen Hochleistungsrechensystemen zu ermöglichen. Dabei wird von dem Benutzer keinerlei Kenntnis über Paralleles Rechnen vorausgesetzt. Dies realisieren wir mit unserem hauseigenen Laufzeitsystems GPI-Space. Dieses ermöglicht es, Algorithmen automatisch und dynamisch zu parallelisieren, wenn sie in Form eines speziellen Graphen, einem sogenannten Petri-Netz, dargestellt werden.

GPI-SPACE