NACHRICHTEN AG

02. August 2024

Neue Methoden für sichere neuronale Netze: Forschung aus Würzburg und München

Wissenschaftler der Uni Würzburg und TU München entwickeln geometrische Methoden zur Robustheit von neuronalen Netzen im Deep Learning.

Wissenschaftler an der Universität Würzburg und der Technischen Universität München setzen sich für die Verbesserung der Robustheit neuronaler Netze ein. Dies geschieht im Rahmen eines Projekts, das von der Deutschen Forschungsgemeinschaft (DFG) gefördert wird. Angesichts der alltäglichen Anwendungen von maschinellem Lernen in Autonomen Fahrzeugen, der medizinischen Bildanalyse und in interaktiven Chatbots wird die Notwendigkeit zunehmend dringlicher, diese Systeme gegen gezielte Angriffe abzusichern.

Die Gefahren gezielter Angriffe verstehen

Neuronale Netze, die für verschiedene Anwendungen eingesetzt werden, zeigen häufig eine Schwachstelle: Sie können durch spezifische Manipulationen von außen gezielt in die Irre geführt werden. Ein Beispiel veranschaulicht dies deutlich: Ein Bild einer Geige kann durch minimale Veränderungen im Pixelbereich von der Software fälschlicherweise als Seelöwe erkannt werden. Solche Fehler sind nicht nur peinlich, sondern können in kritischen Situationen, etwa im Straßenverkehr oder in der medizinischen Diagnostik, katastrophale Folgen haben.

Forschungsverbund und finanzielle

Unterstützung

Der DFG wurde ein neues Forschungsprojekt mit einem Budget von 565.000 Euro genehmigt, wovon mehr als 250.000 Euro an die Julius-Maximilians-Universität Würzburg fließen. Dieses Projekt, welches unter dem Titel „GeoMAR: Geometric Methods for Adversarial Robustness“ bekannt ist, wird in den nächsten drei Jahren erforscht. Die beiden engagierten Wissenschaftler, Professor Leon Bungert und Dr. Leo Schwinn, verfolgen das Ziel, neuronale Netze robuster gegen feindliche Angriffe zu machen.

Robustheit als Schlüssel zur Verlässlichkeit

Robustheit bedeutet in diesem Kontext, dass Systeme sowohl zufälligen als auch gezielten Fehlern gegenüber standhalten sollen. Während moderne Systeme bereits eine beachtliche Widerstandsfähigkeit gegen zufällige Fehler aufweisen, hapert es oft bei gezielten Attacken. Der Forschungsansatz soll neuronale Netze darauf vorbereiten, indem sie in der Ausbildung systematisch mit fehlerhaften und manipulierten Daten konfrontiert werden. Dies soll verhindern, dass Fehler zu gravierenden Problemen führen.

Geometrische Methoden zur Entscheidungsfindung

Ein zentrales Element des Projekts ist der Einsatz geometrischer Methoden. Bei der Klassifizierung von Bildern, etwa bei der Unterscheidung von Hunden und Katzen, muss das neuronale Netz eine Entscheidungsgrenze ziehen. Diese Grenzen sind nicht nur entscheidend für die Genauigkeit des Systems, sondern auch der Hauptangriffspunkt für Manipulationen. Durch das Verständnis dieser Grenzen können neue mathematische Modelle entwickelt werden, die helfen, die Robustheit der Netzwerke zu erhöhen.

Eine zugeschnittene Trainingsmethode

entwickeln

Um die neuronalen Netze optimal zu trainieren, verfolgen Bungert und Schwinn einen innovativen Ansatz: Statt ein einzelnes Netz zu benutzen, trainieren sie zwei unterschiedliche Netzwerke simultan. Eines dieser Netzwerke simuliert die potenziellen Angriffe eines ȣAngreifersȣ. Durch diese Methode hofft das Team, dass das System widerstandsfähiger und einfallsreicher wird als ein menschlicher Angreifer. Ziel ist es, nicht nur die Robustheit zu erhöhen, sondern auch die Effizienz des Trainingsprozesses zu verbessern.

Fazit und Ausblick

Das Projekt ȣGeoMARȣ könnte bahnbrechende Ergebnisse liefern, die nicht nur die Zuverlässigkeit maschineller Lernsysteme erhöhen, sondern auch deren praktische Anwendung in sicherheitskritischen Bereichen fördern. Die Forschungsergebnisse könnten langfristig dazu führen, dass die Technologie des maschinellen Lernens sicherer und effektiver in unserem täglichen Leben eingesetzt werden kann. Der Fokus auf die Robustheit und die Verlässlichkeit dieser Systeme ist von großer Bedeutung, insbesondere in einer Welt, in der autonome Technologien zunehmend in den Vordergrund rücken.

Kontakt

Prof. Dr. Leon Bungert, Professur für Mathematik III (Mathematik des Maschinellen Lernens), Tel: +49 931 31-82849, E-Mail: leon.bungert@uni-wuerzburg.de

Leon Bungerts Webseite

Besuchen Sie uns auf: n-ag.de