

Präzise annotierte Trainingsdaten für CT-Scans

P. Fuchs, Volume Graphics GmbH, Heidelberg

Algorithmen, die auf maschinellem Lernen beruhen, benötigen möglichst viele Trainingsdaten, um ein umfassendes statistisches Abbild der Realität modellieren zu können. Für die zerstörungsfreie Prüfung von Aluminiumgussbauteilen mittels Computertomographie (CT) bedeutet dies, dass hunderte stark Artefakt-behafteter Aufnahmen Voxel für Voxel annotiert werden müssen, um beispielsweise das automatisierte Erkennen von Poren und Lunkern zu trainieren. Dies ist nicht nur sehr aufwendig und fehleranfällig, sondern erfordert auch teures Expertenwissen. Daher sind vor allem für das Training datenhungriger, Deep-Learning-basierter Ansätze zusätzliche Datenquellen wichtig, die über eine einfache Augmentation der Daten – wie drehen, spiegeln, oder das Anpassen von Helligkeit und Kontrast – hinausgehen.

Abhängig von der Zielsetzung lassen sich beispielsweise mit generativen Modellen, die mittels nicht- oder nur schwach annotierter Daten trainiert werden, Trainingsbeispiele für eine einfache Klassifizierung oder grobe Lokalisierung erzeugen. Bei der automatisierten Bauteilprüfung sind jedoch oft abgeleitete Größen, wie Volumen oder Oberfläche, von Bedeutung. Deren Berechnung erfordert eine präzise Segmentierung. Hier bietet die Simulation von CT-Scans die Möglichkeit beliebig genaue Annotationen ohne weiteren Aufwand zu erzeugen. Zudem werden so die physikalischen Effekte der realen Welt zuverlässig wiedergespiegelt. Für das Training neuer Verfahren zur Defekterkennung kann dadurch der Einfluss fehleranfälliger hand-annotierter Daten reduziert werden. Mit einer vollautomatisierten Simulations-Pipeline können beliebig viele Trainingsdaten aus einem künstlich aufgespannten Artefakt-Raum gezogen werden, der die wichtigsten Einflussfaktoren bei der automatisierten Prüfung abgedeckt.