

Evaluationsmetriken für die Bewertung der Bildsegmentierungsmodelle

Erarbeitet von
Dr. Ludmila Himmelspach

Lernziele	1
Inhalt	2
Einstieg.....	2
Dice-Ähnlichkeitskoeffizient	2
Hausdorff-Distanz	4
Ergebnisse der Gewinner der BraTS 2021 Challenge.....	6
Take-Home Message	7
Quellen	8
Disclaimer	8

Lernziele

- Du kannst erklären, welche Evaluationsmetriken für die Bildsegmentierung eingesetzt werden
- Du kannst erklären, wie der Dice-Ähnlichkeitskoeffizient berechnet wird
- Du kannst erklären, wie die Hausdorff-Distanz berechnet wird

Inhalt

Einstieg

Auch ein Bildsegmentierungssystem muss bestimmte Qualitätsanforderungen aufweisen, damit es im laufenden Betrieb eingesetzt werden und zum Beispiel Radiologen bei ihrer täglichen Routine unterstützen kann. In diesem Video erfährst Du, welche Evaluationsmetriken bei der Bewertung der maschinellen Systeme für Bildsegmentierung verwendet werden. Außerdem schauen wir uns an, welche Ergebnisse das Gewinnermodell bei dem BraTS 2021 Wettbewerb erzielt hat.

Dice-Ähnlichkeitskoeffizient

Häufig wird die Bildsegmentierung als eine Klassifikationsaufgabe auf der Pixel- bzw. Voxeleneben aufgefasset. Das heißt, jeder Pixel bzw. Voxel in der Segmentierungsmaske wird entweder der Klasse „zum Objekt gehörend“ oder der Klasse „nicht zum Objekt gehörend“ zugeordnet. Dementsprechend werden für die Bewertung der Bildsegmentierungsmodelle die Evaluationsmetriken für die Bewertung der Bildklassifikationsmodelle angepasst.

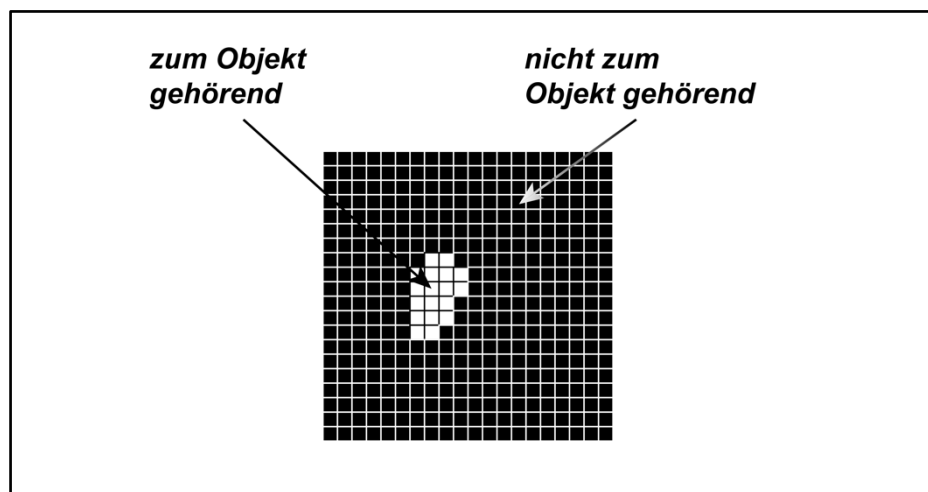


Abbildung 1: Beispiel für eine Segmentierungsmaske.

Eine weit verbreitete Evaluationsmetrik ist dabei der *Dice-Ähnlichkeitskoeffizient (DSC)*, der die Anzahl der richtig erkannten Objektpixel (bzw. Voxel) zur Summe der tatsächlichen und den vorhergesagten Objektpixeln ins Verhältnis setzt.

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

Quelle [1]

Wenn wir annehmen, dass die blau eingefärbte Fläche A der tatsächlichen, also durch die Expert:innen annotierten Fläche des interessierenden Objekts und die rot eingefärbte Fläche B der vorhergesagten Fläche des Objekts entspricht, dann entspricht der Wert des

Dice-Koeffizienten dem Quotienten von der doppelten Überlappung der beiden Flächen, also der lila eingefärbten Fläche, und der Summe der Flächen A und B.

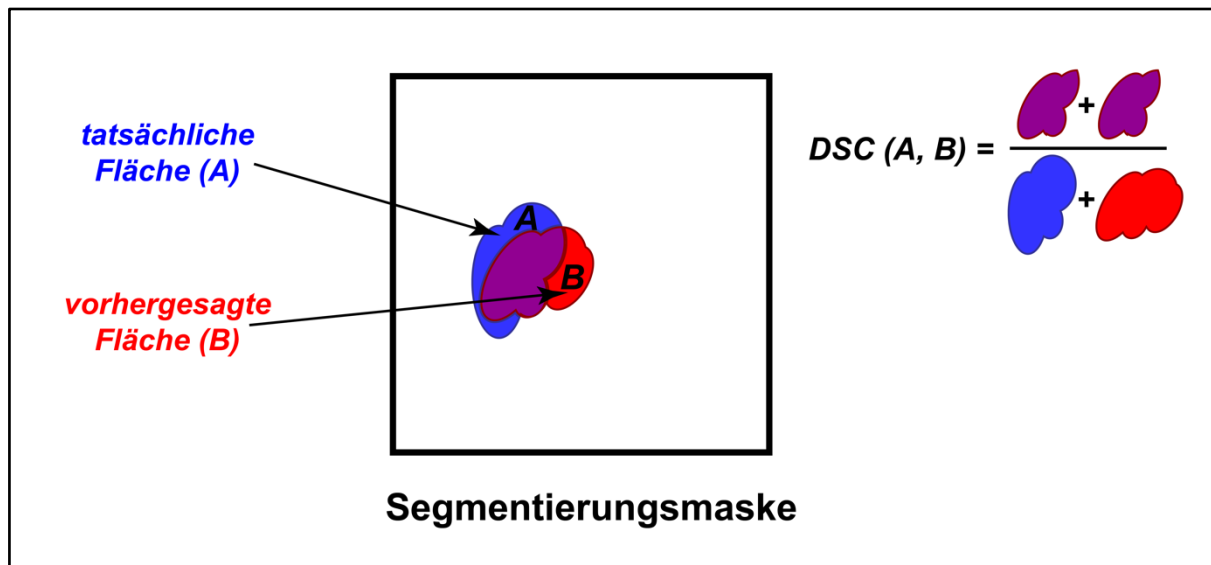


Abbildung 2: Die Berechnung des Dice-Ähnlichkeitskoeffizienten (DSC) zwischen zwei Segmentierungsflächen A und B.

Der Wert des Dice-Ähnlichkeitskoeffizienten liegt im Bereich zwischen 0 und 1 bzw. zwischen 0 und 100 in Prozentangaben. Wenn eine Segmentierungsmethode wie in diesem Beispiel das gesuchte Objekt gar nicht erkennt oder an der falschen Stelle lokalisiert, beträgt der Wert des Dice-Koeffizienten Null. Bei einer präzisen Erkennung des gesuchten Objekts liegt der Wert des Dice-Koeffizienten bei Eins bzw. bei 100 %.

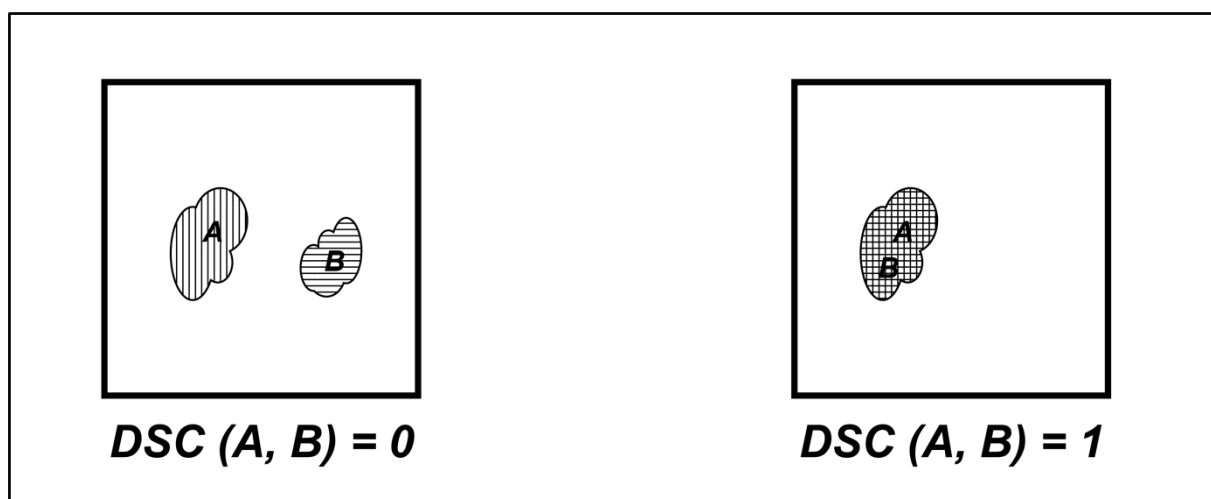


Abbildung 3: Beispiele für die Berechnung des Dice-Ähnlichkeitskoeffizienten bei keiner Übereinstimmung (links) und einer vollständigen Übereinstimmung (rechts) der Segmentierungsflächen A und B.

Hausdorff-Distanz

Eine weitere Evaluationsmetrik für die Bewertung der Bildsegmentierungsmodelle ist die *Hausdorff-Distanzfunktion (HD)*. Die Hausdorff-Distanzfunktion berechnet die maximale Distanz zwischen der annotierten und der durch das System vorhergesagten Segmentierungsfläche.

$$HD(A, B) = \max \left\{ \max_{a \in A} \min_{b \in B} d(a, b), \max_{b \in B} \min_{a \in A} d(a, b) \right\}$$

Quelle [2]

Nehmen wir wieder an, dass die blau eingefärbte Fläche A der durch die Expert:innen annotierten Segmentierungsfläche und die rot eingefärbte Fläche B der vorhergesagten Segmentierungsfläche entspricht. Dann wird zuerst für jedes Pixel bzw. Voxel der Fläche A der Abstand zum nächstgelegenen Pixel bzw. Voxel der Fläche B und umgekehrt für jedes Pixel der Fläche B die Distanz zu seinem nächstgelegenen Pixel der Fläche A berechnet. Der höchste Wert dieser Abstände entspricht dann dem Wert der Hausdorff-Distanz zwischen den Flächen A und B.

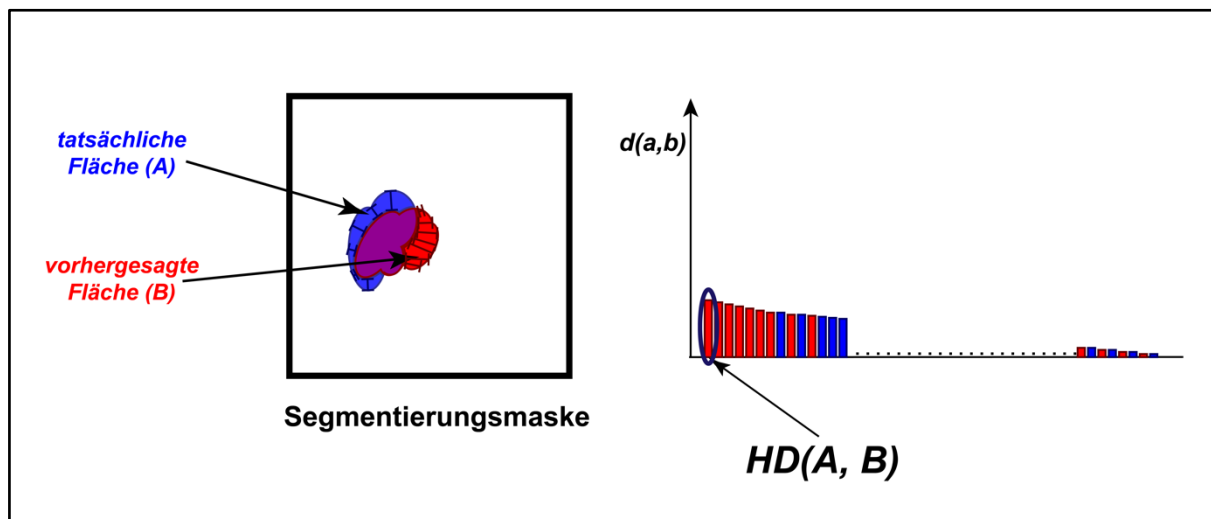


Abbildung 4: Die Berechnung der Hausdorff-Distanz (HD) zwischen zwei Segmentierungsflächen A und B.

Die Hausdorff-Distanzfunktion ist allerdings sehr anfällig für Ausreißer. In diesem Beispiel sehen wir, dass Fläche B bis auf einen kleinen Bereich der Fläche A fast vollständig entspricht. Trotzdem wird die Hausdorff-Distanz zwischen den beiden Flächen durch den Abstand zwischen dem äußersten Randpunkt der Fläche B und dem Rand der Fläche A bestimmt.

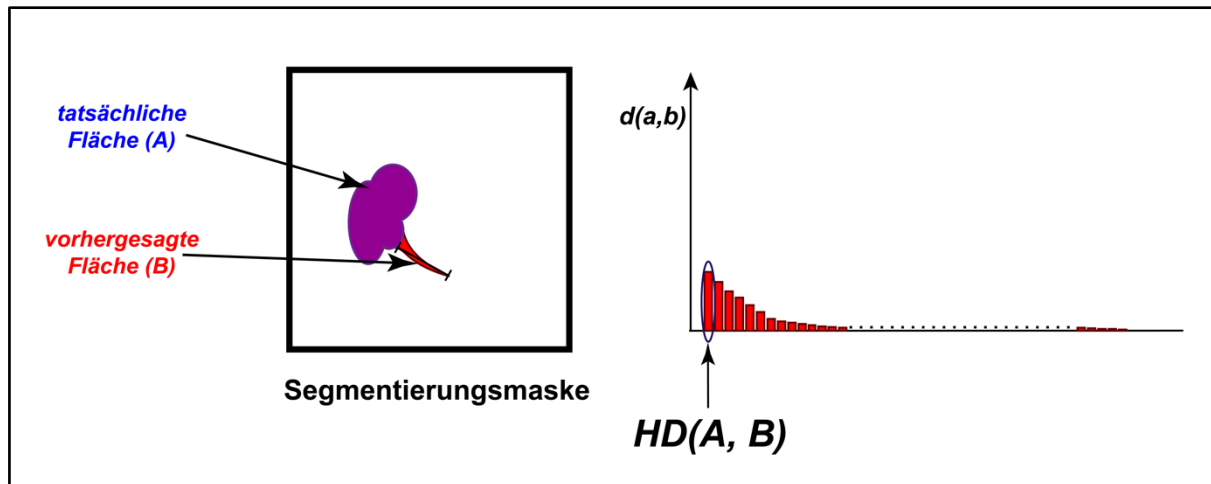


Abbildung 5: Die Berechnung der Hausdorff-Distanz (HD) bei einer fast vollständigen Übereinstimmung der Segmentierungsflächen A und B.

Um den Einfluss von Ausreißern auf die Berechnung der Hausdorff-Distanz abzuschwächen, wird für die Bewertung der Segmentierungsmodelle die 95. Perzentile der Hausdorff-Distanz (HD_{95}) eingesetzt.

Bei der Berechnung der 95. Perzentile der Hausdorff-Distanz zwischen zwei Flächen werden die oberen 5 % der absteigend sortierten Abstände zwischen den Pixeln einer Fläche zu ihren nächstgelegenen Pixeln der anderen Fläche verworfen. Auf diese Weise werden die längsten Distanzen automatisch als Ausreißer erkannt und ignoriert.

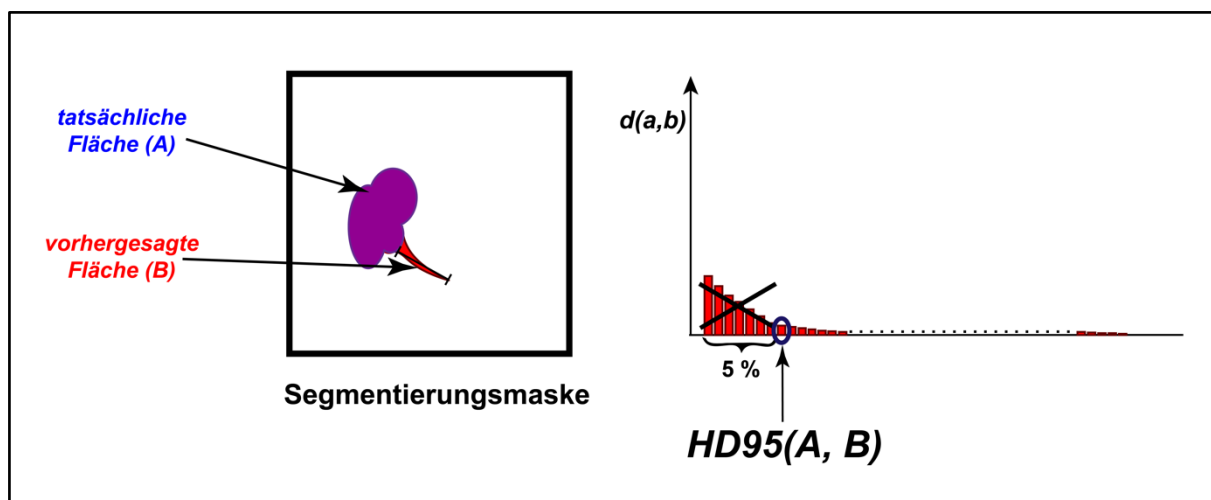


Abbildung 6: Die Berechnung der 95. Perzentile der Hausdorff-Distanz (HD_{95}) zwischen zwei Segmentierungsflächen A und B.

Während der Dice-Ähnlichkeitskoeffizient die Ähnlichkeit zwischen den Volumina zweier Segmentierungen berechnet, berechnet die Hausdorff-Distanzfunktion den Abstand zwischen ihnen.

Ergebnisse der Gewinner der BraTS 2021 Challenge

Die Berechnung der Segmentierungen mittels CNNs auf realen Daten ist sehr rechen- und speicherintensiv, was das Trainieren der neuronalen Netze auf PCs unmöglich macht. Selbst die Speicherkapazitäten der modernen Grafikprozessoren reichen oft nicht aus, um ein U-Net auf großen dreidimensionalen MRT-Aufnahmen ohne Weiteres zu trainieren, weil die zu berechnenden Feature Maps zu groß werden. Deswegen werden die MRT-Bilder sowie die dazugehörigen Segmentierungsmasken oft in kleinere Teile, sogenannte Patches zerlegt und das U-Net auf diesen trainiert. Die berechneten Teilsegmentierungen werden zu einer vollständigen Segmentierungsmaske in einem weiteren Schritt zusammengesetzt.

Diese Strategie haben auch Luu und Park in ihrem Segmentierungsmodell verfolgt, mit dem sie die BraTS 2021 Challenge gewonnen haben. Sie haben noch weitere Veränderungen in ihrem Modell vorgenommen, um es an die Aufgabenstellung anzupassen. Wenn Du Dich für Details des Gewinnermodells interessierst, schau Dir gerne das Paper dazu an. Dort findest Du auch den Link zu ihrem Projekt auf GitHub.

Quelle [3]

Wir sehen uns jetzt aber die Ergebnisse an, die Luu und Park mit ihrem Segmentierungsmodell auf dem BraTS 2021 Datensatz erzielt haben. Wenn Du auf die Werte in der Tabelle schaust, siehst Du, dass die Segmentierungen für drei Regionen der Hirntumore, die durch das Convolutional Neural Network erzeugt wurden, zwar nicht an die Segmentierungen erfahrener Radiolog:innen rankommen, aber sehr nah dran sind. Besonders hervorzuheben sind die Ergebnisse für die Erkennung des gesamten Tumors. Da liegt der durchschnittliche Wert des Dice-Ähnlichkeitskoeffizienten bei fast 93 % und die durchschnittliche Hausdorff-Distanz bei 3,47.

	TC	ET	WT	AVG
DICE	87,81	84,51	92,75	88,36
HD95	7,623	20,73	3,47	10,61

Tabelle 1: Durchschnittswerte für Dice-Ähnlichkeitskoeffizient und 95% Hausdorff-Distanz für das Gewinnermodell der BraTS 2021 Challenge, wie sie von der Online Bewertungsplattform berechnet wurden. TC: Tumorkern, ET: erweiterte Tumorregion, WT: gesamter Tumor (Quelle [3]).

Bei genauer Untersuchung der automatisiert erzeugten Segmentierungen stellten Luu und Park jedoch fest, dass ihr Modell in der Lage war, die Gehirnregionen nahezu fehlerfrei auf MRT-Bildern guter Qualität zu erkennen. Da lagen die Werte für den Dice-Ähnlichkeitskoeffizienten bei über 97 % bzw. 98 %.

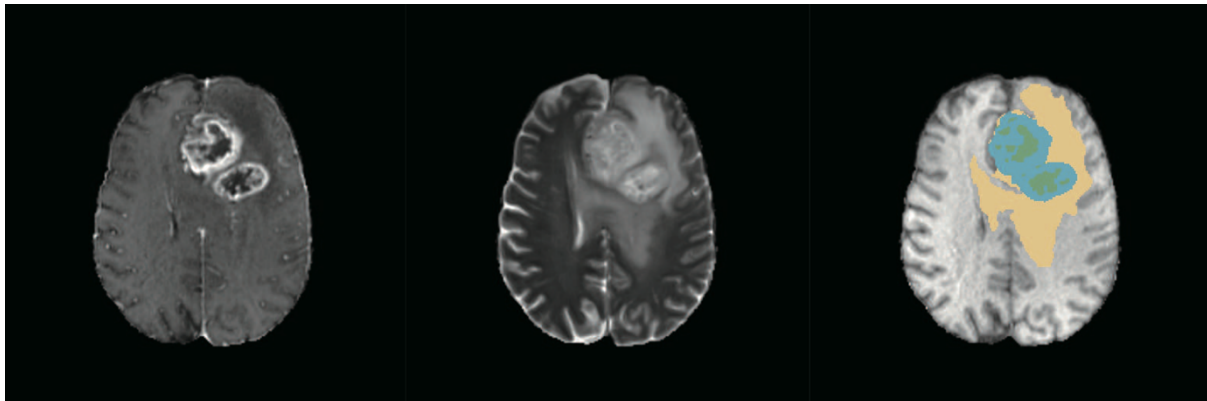


Abbildung 7: Beispiel für eine gute Vorhersage der Segmentierung mit TC: 98.93, ET: 97.78, WT: 98.66 für den Dice-Ähnlichkeitskoeffizient (Quelle [3]).

Auf unscharfen MRT-Aufnahmen, auf denen die Grenzen der Tumorregionen schlecht erkennbar waren, oder auf den Aufnahmen, die sichtbare Artefakte aufwiesen, erkannte ihr System dagegen nur den gesamten Tumor einigermaßen zuverlässig, wobei der Tumorkern weit über die eigentlichen Grenzen geschätzt wurde.

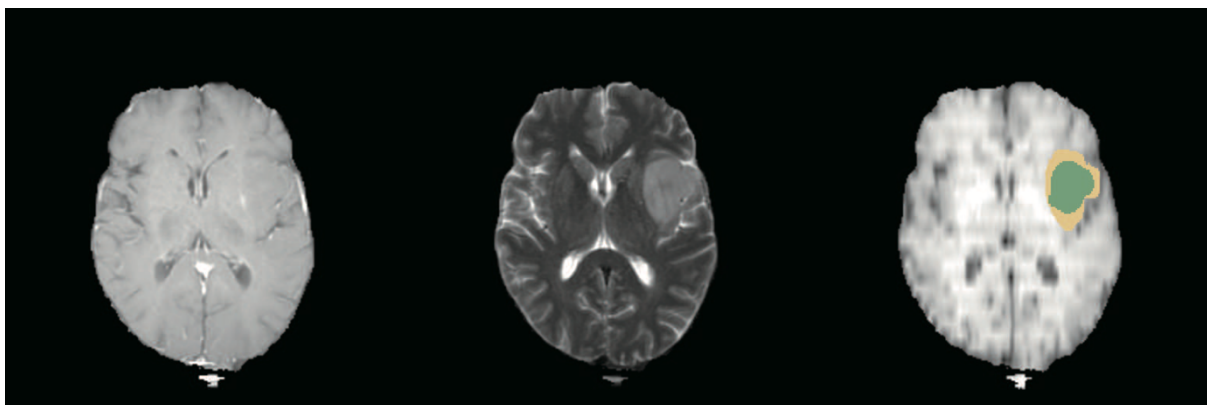


Abbildung 8: Beispiel für eine schlechte Vorhersage der Segmentierung mit TC: 2.76, ET: 0, WT: 92.77 für den Dice-Ähnlichkeitskoeffizient (Quelle [3]).

Dieses Problem deutet nicht unbedingt auf die Schwächen der Architektur des Segmentierungsmodells hin, sondern vielmehr darauf, dass die Datenbasis, mit der das Modell trainiert wurde, nicht ausreichend Randfälle enthielt, mit denen das Modell lernen könnte, sinnvolle Segmentierungen für diese zu erstellen.

Take-Home Message

In diesem Video hast Du zwei Evaluationsmetriken, den Dice-Ähnlichkeitskoeffizienten und die Hausdorff-Distanz, für die Bewertung der Bildsegmentierungsmodelle kennen gelernt. Außerdem hast Du erfahren, welche Segmentierungsergebnisse das Gewinnermodell bei dem BraTS 2021 Wettbewerb erzielt hat.

Quellen

- Quelle [1] Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297-302.
- Quelle [2] Schmitt, I. (2006). *Ähnlichkeitssuche in Multimedia-Datenbanken: Retrieval, Suchalgorithmen und Anfragebehandlung*. Oldenbourg.
- Quelle [3] Luu, H. M., & Park, S. H. (2021, September). Extending nn-UNet for Brain Tumor Segmentation. In *International MICCAI Brainlesion Workshop* (pp. 173-186). Springer International Publishing.

Disclaimer

Transkript zu dem Video „Bildklassifikation und Bildsegmentierung: Evaluationsmetriken für die Bewertung der Bildsegmentierungsmodelle“, Dr. Ludmila Himmelspace.

Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.