

Effect of color space on deep learning algorithm for surgical image classification

Julien Schwerin

Bachelorarbeit • Studiengang Medizininformatik • Fachbereich Informatik und Medien • 13.04.2021

Motivation und Aufgabenstellung

Die Verwendung minimalinvasiver chirurgischer Verfahren ist in modernen Operationssälen nicht mehr nur eine alternative Möglichkeit, sondern meist die erste Wahl. Hierfür werden das Instrumentarium und die Kamera über mehrere kleine Schnitte in den Patienten eingeführt. Neben den vielen Vorteilen für den Patienten, wie beispielsweise einer verbesserten Wundheilung, ergeben sich daraus auch Möglichkeiten der Verarbeitung der Daten, die durch die Verwendung der Kamera gewonnen werden können. Ziel der Arbeit ist die Untersuchung des Einflusses von Farbräumen auf die Performanz von Modellen zur Klassifikation chirurgischer Instrumente. Jene Modelle werden auf den gewonnenen Daten aus chirurgischen Videos trainiert und getestet.

Konzept

Für den Versuch sollen kleine, nicht vortrainierte Modelle zur Bildklassifikation die Gegenwart jeweils eines von vier verschiedenen chirurgischen Instrumenten beurteilen.

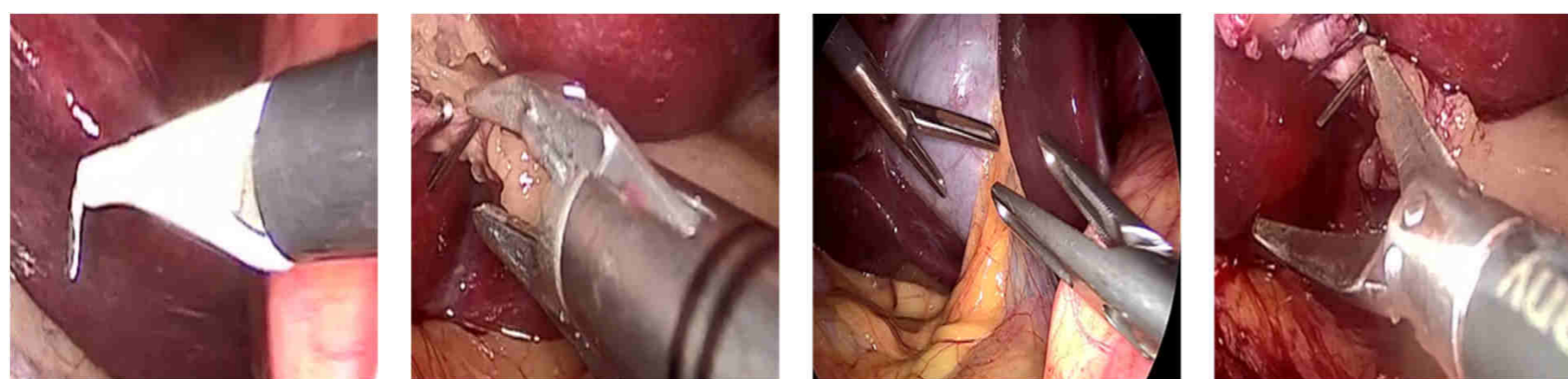


Abb.: Die zu klassifizierenden Instrumente. V. l. n. r.: Haken, Clipsetzer, Greifer und Schere

Als Farbräume für den Versuch werden RGB (Rot-Grün-Blau) und HSV (Hue-Saturation-Value) verwendet. Letzterer wird in zwei Versionen eingesetzt. Einmal werden die Bilder so normalisiert, als wären all seine Kanäle genauso linear aufgebaut wie die Kanäle eines RGB Bildes. Dieser wird als HSVrgb bezeichnet. Die Alternative, mit einer Normalisierung dem zirkulären Aufbau des Hue Kanals entsprechend, wird HSV genannt. Die Normalisierungsfunktionen für diesen Versuch wurden selbst geschrieben.

Es werden Bilder von laparoskopischen Gallenblasenentfernungen des Cholec80-Datensatzes verwendet. Dieser besteht aus 80 Videos und beinhaltet auch die benötigten Label der Instrumentenpräsenz. Je Instrument wird ein eigener, auf das Instrument balancierter Datensatz erzeugt, welcher aus Trainings-, Validations- und Testset besteht.

Auswertung

Für die Auswertung werden die besten Modelle auf einem separaten Testset angewendet und bezüglich ihres erreichten AUROC-Wertes und ihrer Genauigkeit verglichen. Zusätzlich werden Ursachen für die falsch vorhergesagten Bilder mit dem höchsten Confidence-Wert mit Hinblick auf Ursachen für die falschen Vorhersagen betrachtet.

Balancierung der Datensätze

Aus dem Balancieren der Datensätze hinsichtlich der jeweiligen Instrumente resultiert die Tatsache, dass die Modelle nicht direkt auf die Realität übertragen werden können. Das Ziel des Versuchs ist der Vergleich des Einflusses der Farbräume und nicht das Erschaffen von in der Realität anwendbaren Klassifizierern mit möglichst hohen Genauigkeitswerten. Durch das Balancieren entstehen Datensätze unterschiedlicher Größe.

Datensatz Schere	Datensatz Clipsetzer	Datensatz Haken	Datensatz Greifer
6.508 Bilder	11.972 Bilder	162.784 Bilder	163.820 Bilder

Tab. 1: Anzahl der Bilder je Datensatz

Die Modelle

Im Rahmen des Versuchs werden kleine, nicht vortrainierte Modelle verwendet. Nicht vortrainiert aus dem Grund, dass das vorherige Training auf jeweils einem Farbraum statt findet und die Ergebnisse damit verzerrt würden. Für die Verwendung nicht vortrainierter ResNet Architekturen sind die vorhandenen Datensätze wiederum nicht groß genug.

Ergebnisse

Die RGB-Modelle erzielen bei drei von vier Datensätzen die besten Ergebnisse im Hinblick auf Genauigkeit und AUROC. Besonders tun sie sich bei den beiden kleineren Datensätzen Schere und Clipsetzer hervor. Bezogen auf den Greifer schneidet das HSV-Modell vergleichbar gut und beim Haken sogar etwas besser als das RGB Modell ab. Die HSVrgb-Modelle schneiden bei den kleinen Datensätzen und dem Haken ähnlich ab wie das HSV-Modell. Es bleibt jedoch beim Greifer deutlich zurück.

	AUROC			Genauigkeit		
	HSV	RGB	HSVrgb	HSV	RGB	HSVrgb
Schere	0,6565	0,7771	0,7054	0,5880	0,6987	0,6396
Clipsetzer	0,7584	0,8212	0,8050	0,6641	0,7460	0,6469
Haken	0,9855	0,9835	0,9827	0,9505	0,9468	0,9398
Greifer	0,8275	0,8384	0,7688	0,7483	0,7579	0,6963

Tab. 2: Von den Modellen bei verschiedenen Werkzeugen erreichte AUROC- und Genauigkeitswerte. Fett geschrieben sind die besten Werte je Instrument.

Fazit

Sowohl der gewählte Farbraum als auch eine ihm angemessene Normalisierung spielen eine Rolle bei der Betrachtung der Performanz von Modellen zur Klassifizierung chirurgischer Bilder. Im direkten Vergleich der Modelle schneidet der Farbraum RGB besser ab als HSV. Dieses Ergebnis ist ob der Tatsache, dass dieser der aktuelle Standardfarbraum ist, zufriedenstellend. Inwiefern sich die Verwendung anderer Farbräume und Operationen oder größerer Datensätze auf diese Ergebnisse auswirken, lässt sich nicht beurteilen.