



AFCEA Bonn e.V. Studienpreis 2021/2022 Kernthesen der Arbeit

Titel der Arbeit:	An investigation of regression as an avenue to find precision-runtime trade-off for object segmentation
Tag der Einreichung:	15. June 2020
Hochschule:	H-BRS
Name des Verfassers:	Arun Rajendra Prabhu
Betreuender Professor:	Prof. Dr. Paul G. Plöger

Kurze Beschreibung (1 Seite !) der Kernthesen. Was ist die Quintessenz der Arbeit?

Die Arbeit behandelt ein Thema des Maschinellen Lernens mit Anwendung auf den Bereich Autonomes Fahren. Typische Aufgaben dort sind z.B. die Objekt-Erkennung (OE) und -Klassifikation oder das Finden und Maskieren von Objekt-Instanzen (MOI). Dabei gilt es z.B. parkende Autos auf dem Seitenstreifen als Objekt AUTO zu erkennen und auszublenken. In der aktuellen Forschung gibt es so eine Vielzahl von Algorithmen und zusätzlich dazu annotierte Standard-Datensätze, mit denen neue Algorithmen validiert werden können (siehe Tab. 3.1).

Die vorliegende Arbeit analysiert einige der dieser existierenden MOI-Lösungen im Detail, deckt Schwachstellen auf und implementiert insgesamt drei Verbesserungen in Hinblick auf verkürzte Laufzeiten und die erreichte Genauigkeit (mean average precision, mAP).

Erstens wird die MOI-Aufgabe - um die Inferenzzeit zu verkürzen - nicht mehr als ein Klassifikationsproblem auf Pixel-ebene betrachtet, sondern als eine geometrische Regressionsaufgabe, d.h. man finde eine Figur, die ein Objekt möglichst exakt umschließt. Um die Genauigkeit zu verbessern, schlägt der Autor hier statt der üblichen umschreibenden Bounding Box (BB) eine wesentlich genauere achteckige Maske vor (Octagonal Mask, OCM), denn BBs können für konvexe Objekte relativ viele falsch-positive Pixel vereinnahmen. Grundlage für die Verbesserung bildet eine geschickte Modifikation einer BB-basierten Kostenfunktion, um damit die OCM Kosten zu modellieren. Die Implementation fußt dabei auf einem tiefen Neuronalen Netz vom Typ SqueezeDet+, das nach der Anpassung SqueezeDetOcta genannt wird. Die Verwendung von OCMs zieht nun nach sich, dass auch alle üblichen Erfolgsmetriken wie z.B. die mAP in Hinblick auf die Achtecke überarbeitet werden muss, da diese nur für achsen-parallelen BBn gültig sind. Zum Training und zur Verifikation steht der Cityspace Datensatz zur Verfügung, der pixelgenaue Objektmasken als Referenz bereitstellt.

Zweitens gehört SqueezeDet+ zu den ankerbasierten Verfahren. Anker sind BB-Hypothesen unterschiedlicher Größe und Position, die über das gesamte Bild verteilt werden und die bezeichnen, wo und in welcher Größe MOIs zu erwarten sind. Hier fiel dem Autor auf, dass Unstimmigkeiten bestehen zwischen dem Ankerdesign und den beobachteten Masken für die Referenz-Objekte. Die Breiten und Höhen der Anker wurden für andere State-of-the-Art Ankerverfahren entweder handverlesen oder durch Clusterbildung berechnet. Hier kann der Autor zeigen, dass durch einfaches Logarithmieren der Breiten und Höhen eine wesentlich gleichmäßigere Abdeckung der möglichen Ankerhypothesen erreichen lässt, und so die mAP signifikant verbessert werden kann.

Drittens und letztens ist dem Autor aufgefallen, dass Objekte, die am Rand anhaften (Teilinstanzen), die Trefferrate schon beim Training negativ beeinflussen. Denn solche Objektinstanzen entsprechen eigentlich nicht der Lernaufgabe, da sie künstlich durch den Rand beschnitten wurden und dies Phänomen nicht korrekt durch eine dort endende BB beschrieben wird. Der Autor implementiert daraufhin eine Methode, diese Objekt-Instanzen automatisch erkennen kann und sie so während des Trainings ignoriert werden. Durch dieses selektive Herausfiltern von verfälschten Informationen werden nur noch die unverfälschten Informationen verwendet, was sich wieder signifikant auf die mAP auswirkt. Diese Problemstellung ist meines Wissens genuin neu und wurde bisher in der Literatur übersehen.

Ich halte die Arbeit für preiswürdig aus folgenden Gründen: obwohl sie mit ihren vielen kleinteiligen evolutionäre Änderungen für keinen „großen“ Durchbruch sorgt, ist das mittlerweile typisch für das Feld, weil Durchbrüche nur noch mit riesen Teams mit riesen Rechenzeiten möglich sind. Dadurch unterliegt die Arbeit ständig der Gefahr der Fragmentierung. Deshalb ist es so herausragend, wie die Arbeit mittels ihrer methodisch sehr strengen, deduktiven Arbeitsweise "vorlebt", wie systematisches wissenschaftliches Arbeiten funktioniert: dies hält die Argumentationslinie auf Kurs. Es wird eine klare Nomenklatur eingeführt (i.e. in allen abgerundeten Text-Boxen). Diese begründen schon die Designscheidungen bei der Vorauswahl und dem Setup. Weiter bezeichnen sie alle Hypothesen, die mit entsprechenden Experimenten falsifiziert oder verifiziert werden (ausgedrückt durch grüne oder rote Rahmenfarbe). Bei unklarem Versuchsausgängen / unbeantwortbaren Fragen werden neue verfeinerte Hypothesen aufgestellt oder weitere Experimente gemacht oder Fragen verfeinert, falls Hypothesen so nicht beantwortbar sind. Diese Vorgehensweise ist mustergültig. Die Arbeit wurde als Technical Report der HBSR veröffentlicht (<https://doi.org/10.18418/978-3-96043-086-5>). Sie dient seither als Referenz für Masterarbeiten in denen multi-faktorielle Design-Räume untersucht werden sollen. Darüber hinaus führt die Masterarbeit eine neue Darstellungsform (Parallel Flow Diagrams) ein, die die Ergebnisse der Design Space Exploration sehr nachvollziehbar und hübsch darstellen.

