

Depth Estimation with Deep Learning from Video Images with LIDAR Ground Truth

Studiengang: MAS Data Science

Die Aufgabe der Entfernungsabschätzung von Objekten in zweidimensionalen Indoor-Fotos soll mit Hilfe einer ausgewählten Deep Learning Architektur umgesetzt und analysiert werden. Für die Bereitstellung der notwendigen Trainingsdaten dient eine Azure Kinect DK Light Detection and Ranging (LIDAR) Kamera.

Einleitung

Bei vielen foto- bzw. videoverarbeitenden Anwendungsfällen sind Entfernungsinformationen von besonderer Bedeutung. Oft ist es nicht ausreichend, Objekte oder Bereiche auf Fotos, bzw. in Videos identifizieren zu können, und oft ist es auch erforderlich, die genaue Entfernung zu diesen Objekten zu wissen. Beispiele für diese Anwendung sind nachträgliche Tiefenunschärfenkorrekturen, der Parallaxen Effekt oder der 3D Ken Burns Effekt, welcher im verlinkten Video (Quelle: Simon Niklaus, Long Mai, Jimei Yang and Feng Liu - 3D Ken Burns Effect from a Single Image) an zweidimensionalen Bildern, ohne zusätzliche Informationen, angewendet worden ist. Diese Master-Thesis beschäftigt sich mit der notwendigen Grundlage, um solche Effekte erst zu ermöglichen.

Inhalt

Neben den theoretischen Grundlagen von Künstlicher Intelligenz, Convolutional Neural Networks sowie die für die Vergleichbarkeit und Qualitätsaussage notwendigen Evaluierungsmetriken, wird vor allem das Thema Ground Truth Generierung und die Auswahl einer geeigneten Deep Learning Architektur für die Aufgabe der monokularen Tiefenerkennung (mono-

cular Depth Estimation) näher beleuchtet. Neben der detaillierten Erläuterung der gewählten AdaBins-Architektur von Shariq Farooq Bhat, Ibraheem Alhashim und Peter Wonka sind vor allem die mit den zur Verfügung stehenden Datensets, NYUv2 und „meine Aufnahmen“ erzielten Ergebnisse interessant. Abbildung 1 zeigt zwei Beispielbilder. Links, die Quell-RGB-Bilder, gefolgt von den vorhergesagten Bildern, auf Basis des mit den selbst-generierten Daten neu-trainierten Modells. Daneben die vorhergesagten Bilder, auf Basis des NYUv2 Datensets vor- bzw. neu-trainierten Modells, sowie die selbst-generierten Ground Truth Depth-Images.



Christian Rieser

Schlussfolgerung / Fazit

Im finalen Kapitel werden die Ergebnisse diskutiert und mit Verbesserungsmöglichkeiten und weiteren Schritten vervollständigt. Im Fazit wird nochmals auf die Wichtigkeit der Quelldatenqualität für Deep Learning Modelle hingewiesen.

Die Master-Thesis ist für jeden interessant, der sich mit den Themen der Ground Truth Erstellung und der monokularen Tiefenerkennung von RGB-Bildern beschäftigt, oder in ähnlichen Bereichen tätig wird.

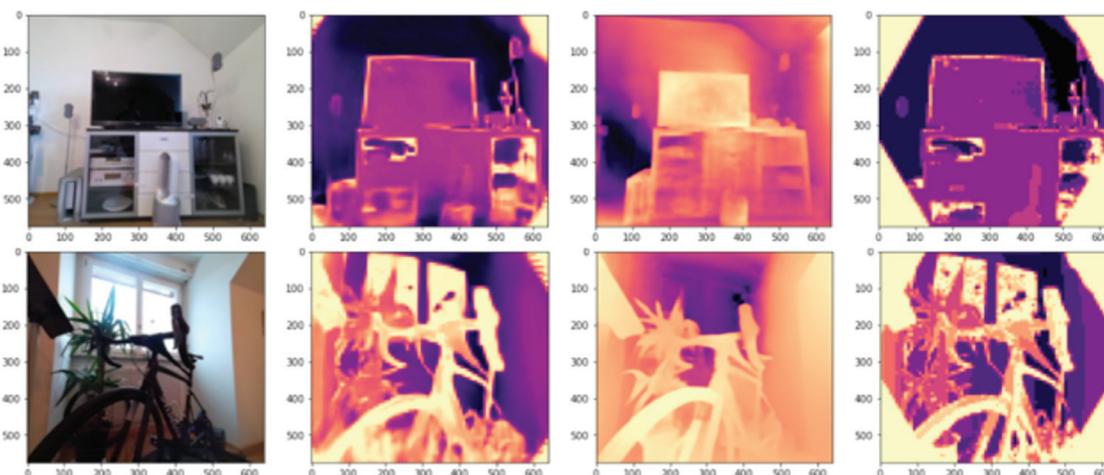


Abbildung 1 - Zwei vorhergesagte Beispielbilder inkl. RGB- und Ground Truth Depth-Image