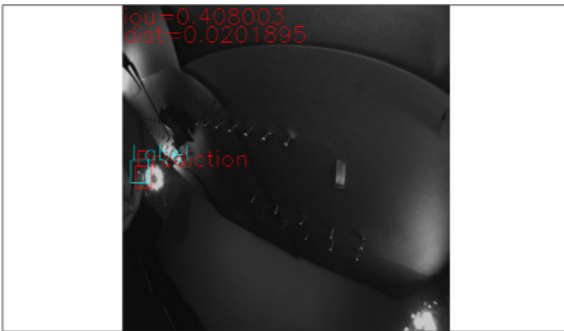
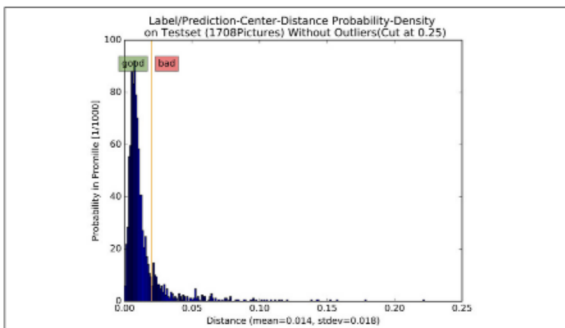


Student	Heinz Hofmann
Examinator	Prof. Dr. Guido Schuster
Themengebiet	Sensor, Actuator and Communication Systems

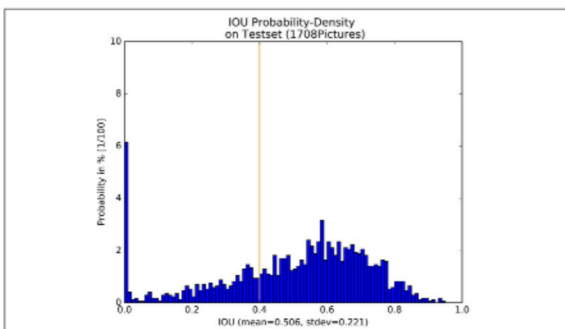
## Yolo auf Finger



Beispiel eines Bildes mit Label(türkis) und Prediction(rot)



Histogramm der Distanz zwischen Label und Prediction



Histogramm der IOU zwischen Label und Prediction

**Aufgabenstellung:** Das Ziel dieser Projektarbeit war es, herauszufinden, ob Yolo geeignet wäre, die Fingerspitzen einer Hand in einem Bild zu klassifizieren und genau zu detektieren. Die Vorgaben, was die Genauigkeit betreffen liegen bei 0.1mm.

Yolo ist eine Möglichkeit, um mittels Deep-Learning Objekte in einem Bild zu klassifizieren und gleichzeitig deren genaue Position zu detektieren. Daher auch der Ausdruck Yolo (You only look once). Yolo wurde als Konzept gewählt, weil es in diesem Bereich dem aktuellen Stand der Technik entspricht. Gerade die Geschwindigkeit dieses Netzwerks wird als extrem hoch angepriesen (bis zu 45fps). Diese Geschwindigkeit ist für die letztendliche Anwendung von hoher Wichtigkeit, weil es sich letzten Endes um eine Echtzeitanwendung handeln soll.

**Vorgehen:** Mithilfe der Apparatur und Software von Tabea Mendez wurden zuerst Daten generiert. Um diesen Aufwand klein zu halten, wurden nur Daten vom rechten Zeigefinger generiert. Gleichzeitig wurde in Tensorflow die Architektur von Yolo nach gebaut. Dies wäre nur begrenzt nötig gewesen, da fertige Architekturen in Keras oder Darknet online zur Verfügung stehen würden. Um aber einen Lerneffekt im erstellen von Neuronalen Netzwerken zu erzielen, wurde trotzdem alles von Grund auf selber aufgebaut. Rund um die Kernarchitektur von Yolo wurde das Datenhandling, die Kostenfunktionen aber auch sämtliche Validierungen und Tests zweimal erstellt. Einmal für das Pretraining der Kerngewichte auf dem ImageNet Klassifizierungsdatenset und einmal für das "echte" Training auf den selber generierten Daten. Sobald dies alles aufgebaut und lauffähig war, wurde noch so viel wie möglich experimentiert, um herauszufinden, mit welchen Änderungen und Einstellungen das Lernresultat noch optimiert werden könnte.

**Fazit:** Das Pretraining und auch das Training hatten seine Tücken, weil das originale Yolo-Netzwerk extrem gross ist, und entsprechend nahezu den ganzen RAM-Speicher einer GPU benötigte, wodurch nur noch begrenzt Platz für Daten übrigblieb. Diese Probleme konnten einigermaßen umgangen werden, hatten jedoch zur Folge, dass die Bilder von 1280x960 auf 448x448 verkleinert werden mussten, um das Netzwerk zum laufen zu bringen. Dies hatte zur Folge, dass ein Pixel bereits bis zu 1,5mm entsprechen konnte. (Dies sollte Yolo theoretisch nicht daran hindern genauere Aussagen über die Position des Fingerspitzes zu machen.) Trotzdem wurde mit rund 84% der Predictions nur eine Genauigkeit von 15mm erreicht, was in etwa 10 Pixeln entsprach. Mit diesem Ergebnis wurde zwar das Ziel der Aufgabenstellung (0.1mm) um Faktor 100 verpasst, allerdings in 84% der Fälle Predictions gemacht, welche aus subjektiver menschlicher Sicht "gut" sind. Dies ist ein einigermaßen Erstaunliches Resultat, wenn man bedenkt, dass man zum Trainieren nur rund 18'000 Bilder verwendet hatte. Es ist anzunehmen, dass mit einer Verbesserung der Datengewinnung und entsprechend viel mehr Daten in naher Zukunft mit diesem oder einem ähnlichen Konzept eine Genauigkeit von bis zu 1mm erreicht werden können sollte.