

## Rättvisare maskininläring inom datorseende

Datorseendet har inom vissa områden redan passerat människans förmåga. Vi upptäcker dock allt fler exempel på att fördomar har inprogrammerats även i de mest framgångsrika maskininlärningsmodellerna. Forskning visar att många modeller för exempelvis ansiktsigenkänning (face detection, face landmarks detection) presterar bättre på människor som uppfattas vita. I ordrepresentationer inom datorlingvistik har det även påvisats stereotypa könsuppfattningar som att kopplingen mellan exempelvis orden man och ingenjör kan vara signifikant starkare än den mellan kvinna och ingenjör. Dessa problem är ofta ett resultat av obalanserad data och att det ännu inte finns en standardiserad metod bland forskare inom maskininläring för att upptäcka, mäta och minimera bias i nya modeller.

Ny forskning inom området visar att modifieringar i modellens kostnadsfunktion kan vara en framtida metod för att minska bias. Metoder för att hantera skev data finns även inom djup maskininläring för medicinska tillämpningsområden och skulle kanske kunna överföras till detta problem.

Målet med kandidatarbetet är att studera en modell och tillämpning som har en tydligt påvisad bias på ett känt dataset för exempelvis ansiktsigenkänning. Med utgångspunkt i tidigare forskning vill vi sedan jämföra existerande och utveckla modifieringar till modellens kostnadsfunktion som minskar modellens bias. Utöver detta kan även data-augmentering eller specifika samplings- och träningsmetoder studeras med samma mål. Större beräkningar kan göras på beräkningsklustret Alvis, <https://www.c3se.chalmers.se/about/Alvis/>.

Artiklar relaterade till ämnet:

Investigating Bias in Deep Face Analysis: The KANFace Dataset and Empirical  
Markos Georgopoulos and Yannis Panagakis and Maja Pantic, 2020  
<https://arxiv.org/abs/2005.07302>

Racial Faces in-the-Wild: Reducing Racial Bias by Information Maximization Adaptation  
Network, Mei Wang and Weihong Deng and Jiani Hu and Xunqiang Tao and Yaohai Huang,  
2019  
<https://arxiv.org/abs/1812.00194>

**Handledare** Mats Granath, mats.granath@physics.gu.se

**Studentgrupper** F och TM

**Gruppstorlek** 3-6 studenter. Kan inte dubbleras

**Initiativtagare och kontaktperson** Sanna Persson, pesanna@student.chalmers.se

**Förkunskapskrav**

Kunskap inom maskininläring motsvarande Deep Learning Specialization av Andrew Ng på Coursera är fördelaktigt samt tillräckliga programmeringskunskaper för att kunna lära sig bibliotek för djup maskininläring som PyTorch/Tensorflow i Python.