

Tolkningsbara Representationer av Ansikten med Artificiell Intelligens

Författare: Erik Sandström

Populärvetenskaplig sammanfattning av:

Latent Space Growing of Generative Adversarial Networks

Maskininlärning har de senaste åren bidragit till starten på ett omfattande paradigmskifte mot ett smartare och mer uppkopplat samhälle. Det här examensarbetet bidrar till den utvecklingen genom att öka förståelsen kring hur bilder på ansikten kan representeras så att ansiktsdragen hos personen blir naturligt beskrivna i representationen. Resultatet är ett system som kan generera verklighetstroga bilder på ansikten och modifiera dessa på ett kontrollerat sätt.

I det här arbetet presenteras en så kallad generativ modell, som kan generera verklighetstroga bilder på ansikten på ett nytt och mer flexibelt sätt. Ett stort problem med dagens generativa modeller är att det ofta är svårt att skapa ansikten där användaren kan specificera till exempel hårfärg, kön, ögonfärg, ansiktsuttryck etcetera. Istället skapas ett slumpmässigt, men verklighetstroget ansikte, som är svårt att modifiera. Modellen som konstruerats för det här examensarbetet gör det lättare att modifiera redan skapade ansikten. Andra fördelar med den nyframtagna modellen är att den lär sig snabbare än existerande modeller och ger bilder med åtminstone lika god kvalitet som de bästa modellerna tillgängliga idag.

För att generera ett ansikte används en mycket komplicerad matematisk funktion som bekvämt nog kallas för generator. Generatorn behöver en insignal för att ge ett ansikte som utsignal och det är insignalen som vi kallar för representationen av ansiktet. Insignalen eller representationen består av flera siffror. Varje tal påverkar på något sätt utseendet av ansiktet, men det är väldigt svårt att relatera varje enskild siffra till ett separat ansiktsdrag, till exempel hudfärg, form på ansiktet och så vidare. Målet med modellen som konstruerats i det här arbetet är att starkare koppla insignalens effekt till specifika ansiktsdrag. Genom att stegvis öka antalet insignalssiffror till generatorn när modellen lär sig så påverkar siffror som tränats tidigt grova ansiktsdrag som till exempel kön, hårfärg, frisyr och hudfärg medan senare tränade siffror påverkar finare detaljer som till exempel ögonfärg och -form samt detaljer i ansiktsuttrycket. Ett exempel ges i Figur 1a och 1b där varje rad beskriver hur ett

visst ansikte förändras då en specifik siffra i insignalen till generatorn ändrar värde. I Figur 1a sker stora förändringar, eftersom en siffra som tränades tidigt ändras. I Figur 1b ändras mer eller mindre endast hur mycket ögonen är öppna, eftersom en siffra som tränades sent ändras.

Den typen av generativ modell som används i det här arbetet kallas på engelska för "Generative Adversarial Network". Även om namnet låter aningen skräckinjagande, eller kanske bara häftigt, så är intuitionen bakom modellen väldigt simpel. Generatorns uppdrag är redan beskrivet. Den behöver dock tränas för att generera verklighetstroga ansikten. Till dess hjälp finns en annan funktion, kallad diskriminator, som försöker se skillnad på riktiga ansikten och ansikten skapade av generatorn. Generatorns uppdrag är därför att lura diskriminatoren medan diskriminatoren agerar polis och försöker upptäcka de "förfalskade" ansiktena. Diskriminatoren ger feedback till generatorn om huruvida den lyckas skapa bra bilder och när diskriminatoren inte längre kan se skillnad på riktiga ansikten och förfalskade ansikten har generatorn tränats klart.

I detta arbete har enbart bilder på ansikten studerats, men i ett större sammanhang kan resultaten från detta arbete ses som en del i att producera bra representationer av godtycklig data. Representationer är mycket viktiga. Till exempel är en bild i en dator inte en bild så som vi människor tänker på den, utan bara en stor tabell med pixelvärden. Tabellrepresentationen är användbar för att visa bilden på datorn, men inte för att hjälpa datorn att "förstå" innehållet i bilden. Med en annan representation kan samma bild beskrivas, men med tydligare semantisk koppling till motivet. Att definiera vad som gör en "bra" representation är svårt eftersom det är problemspecifikt. För att till exempel klassificera en bild på ett djur kan en relevant representation vara till exempel hur många ben djuret har, om djuret har en mönstrad päls, om det har en svans, om det har klor etcetera. Om problemet är att räkna antalet röda pixlar i bilden hjälper däremot inte den beskrivna representationen. För naturliga problem som till exempel ansiktsigenkänning eller objektsigenkänning hjälper dock en representation som är semantiskt tolkningsbar och här kan detta arbete visa på en potentiell väg för framtida forskning inom maskininlärning och datorrepresentationer av bilder.



(a)



(b)

Figure 1: Varje rad beskriver hur ett ansikte ändras då signalen till generatorn ändras. (a) Stora förändringar då en siffra i signalen som lades till tidigt under inläringen ändras. (b) Små förändringar då en siffra i signalen som lades till sent under inläringen ändras.