

EXAMENSARBETE The Effect of Synthetic Data in Training of

Deep Neural Networks for Object Detection in Cityscapes

STUDENTER Arvid Mildner, Tony Liu**HANDLEDARE** Pierre Nugues (LTH); Martin Ljungqvist, Otto Nordander (Axis Communications)**EXAMINATOR** Jörn Janneck (LTH)

Träning av objekt-detektor med hjälp av syntetiska bilder i stadsmiljö

POPULÄRVETENSKAPLIG SAMMANFATTNING **Arvid Mildner, Tony Liu**

AI-modeller som används för objekt-detektion är i behov av stora mängder exempelbilder på objekt såsom bilar och personer. Det är dyrt att både samla in och annotera dessa bilder. Därför skulle ett system som kan tränas på datorgenererade bilder kraftigt minska kostnaden att träna modeller i nya miljöer.

Man skulle kunna tänka sig att det inte är så stor skillnad på en riktig bild och en detaljerad syntetisk bild. För det mänskliga ögat verkar det i alla fall inte svårare att urskilja en bil i ett datorspel än vad det är om man tittar på en riktig bild. Den typ av modell som vi har experimenterat med är ett så kallat neuralt nätverk som på en konceptuell nivå fungerar ungefär som man tror att vår egen hjärna behandlar synintryck.



(a) Riktig bild från Berkeley (b) Syntetisk bild från datorspelet GTA V.

Tyvärr visar det sig att det inte är så enkelt - en modell som enbart har sett syntetiska bilder har svårt att få god precision när den tillämpas på riktiga bilder. Vår idé var att försöka förstå varför det ligger till på det sättet och se om vi istället kan göra något för att använda så få riktiga bilder som möjligt tillsammans med de syntetiska bilderna. Först experimenterade vi med något som kallas för *transfer learning* för att nå en så god precision som möjligt given en viss mängd riktig data. Grundidén med den här metoden är att ett nätverk som har sett syntetisk data enbart behöver en liten skjuts i rätt riktning genom att få se riktig data mot slutet av träningen. Då ska den ha lärt sig i stora drag vad den ska kolla efter av den första träningen med syntetisk data. Här visade vi att syntetisk data kan vara mycket värdefullt om man

har få riktiga bilder men har avtagande effekt på precisionen när mer riktig data introduceras.

Därefter försökte vi dyka in i nätverkets inre struktur för att försöka hitta vad som egentligen skiljer en modell som tränats med hjälp av syntetisk data och en annan som tränats med riktig data. Här presenterar vi olika matematiska metoder för att studera likheten mellan de interna representationerna i nätverket som leder fram till de prediktioner som görs. En allmän uppfattning inom forskningsområdet är att de olika interna lagrena i nätverket får olika uppgifter som sträcker sig från att detektera konturer till att detektera faktiska saker som hjul och bilfönster. Detta gör att nätverket slutligen kan hitta hela bilen allt eftersom den ursprungliga bilden filtreras genom hela nätverkskroppen. Idén var att vi skulle kunna isolera delar av nätverket som verkar göra ungefär samma sak oberoende av om de enbart sett syntetisk data eller riktig data.

Det visade sig vara svårare än vi trodde att se något anmärkningsvärt från den här analysen. Mycket av de slutsatser vi kan dra hamnar lite i skymundan bakom detaljer som har att göra med modellens arkitektur snarare än det faktum att modellerna sett olika typer av data. Trots detta ger vår analys indikationer på att en modell som tränats på både syntetisk och riktig data kan hitta nya objekt-detektioner och generaliserar bättre till nya miljöer jämfört med en modell tränad på enbart riktig data. Resultatet av detta är en sorts granulär lupp som skulle kunna användas vid vidare forskning på hur ett nätverk kan bli mer effektivt med hjälp av syntetisk data.