Centre for Mathematical Sciences | LUNDS TEKNISKA HÖGSKOLA | PRESENTERAD 2023-05-31

EXAMENSARBETE Detecting Images Outside Training Distribution for Fingerprint Spoof Detection
STUDENT Daniel Holmkvist
HANDLEDARE Ida Arvidsson (LTH) och Ulf Homlstedt (Precise Biometrics)
EXAMINATOR Kalle Åström (LTH)

Detektion av fingeravtrycksförfalskning utanför träningsdatan

POPULÄRVETENSKAPLIG SAMMANFATTNING **Daniel Holmkvist**

Moderna maskininlärningalgoritmer har ofta svårt att hantera data annorlunda till det den tidigare har sett. Detta arbete undersöker möjligheten att detektera huruvida en ny bild algoritmen får se är olik träningsdatan för att förbättra fingeravtrycksdetektion som utförs på företaget Precise Biometrics.

Maskininlärningsalgoritmer används idag i många fält, och har visat sig väldigt användbara till allt från självkörande bilar till chatbottar. Algoritmerna använder data för att skapa modeller till att förutspå framtiden. De är dock inte utan problem, utan kan ha svårt att hantera data olik den det tidigare har sett. När nätverket får sådana bilder kan resultatet ofta bli konstigt och oförväntat, något som kan vara väldigt negativt för säkerhetskritiska applikationer så som fingeravtrycksdetektion. Men det finns metoder för att upptäcka och flagga dessa bilder, några av vilka jag i detta examensarbete undersökt.

Datan jag har använt mig av för att utföra tester kan primärt delas upp i två delar. Första gruppen består av bilder som är väsentligt skilda ifrån träningsdatan. Dessa är inte fingeravtryck överhuvudtaget, utan istället exempelvis siffror eller byxor. Den andra gruppen är svårare och består av bilder som är liknande träningsdatan, det vill säga fejkade fingeravtryck som nätverket inte tidigare har sett. I literaturen används metoderna primärt på bilder som är väsentligt skillda träningsdatan, medan jag här undersökt hur bra metoderna fungerar på denna "svårare" data, och vad som kan förbättra dem i denna kontext.

Jag har använt mig av nya metoder som i forskningen visat bra resultat i studier, både i riktighet och snabbhet. Metoderna har även gemensamt att de inte kräver en förändring av nätverkets uppbyggnad, eller kräver att hela nätverket tränas om. Metoderna har förbättras specifikt på det här problemet, både genom att förbehandla datan samt genom lämpliga parameterval. Som nätverk användes kraftfulla fingeravtryckssklassifierare från Precise Biometrics vilka metoderna integrerades med för att ta hand om ny, skild data.

Resultatet visar att metoderna fungerar väldigt bra på data som är väsentligt skild ifrån träningsdatan, där dessa metoder betydligt förbättrar sannolikheten att algoritmen ger en önskvärd respons. På sådan data kan man uppnå rätt resultat i över 99% av fallen med den bästa metoden jag undersökte. Det är dock som förväntat mycket svårare på data som är mer lik fingeravtryck. Här kan man inte förvänta sig i närheten av den typen av resultat, men beroende på kvaliteten av de falska fingeravtrycken kan man ändå prestera en bra bit över slump. För att försöka få ett bättre resultat skulle man kunna ta till många olika strategier. Till exempel kan man använda fejkade fingeravtryck nätverket inte tidigare sett för att träna detektorn, eller utforska andra metoder.