

# Vad krävs för att förbättra ett neuralt nätverk att känna igen människors rörelser från nya vinklar?

Filip Cederquist  
Gustav Harrysson

**Att kunna tolka en persons rörelsemönster är någonting människor tar för givet men för en dator är det mer komplicerat. Vårt resultat tyder på att det är möjligt att med relativt lite ny information, avsevärt förbättra datorns förmåga att känna igen rörelser från nya vinklar.**

Axis communications är ett företag som håller på med videoövervakning. De flesta videos som Axis behandlar är från ett säkerhetskameraperspektiv vilket är ett område med mycket potential för videoanalys. För nuvarande är det stort fokus på vad som finns i bilden, nästa steg är att undersöka vad det är som händer. Ett problem idag är att datorer främst undersöker en video genom att titta på varje stillbild separat. Genom att utveckla denna analys till att titta på sekvenser är det möjligt att få en mer kontextuell uppfattning av vad som försigår. Med information om mer specifika händelser kan videoövervakning automatiseras och minska behovet av en iakttagande människa. För att lösa dessa komplicerade problem används en struktur som kallas neurala nätverk. Dess design påminner om människans hjärna, det finns miljontals neuroner som aktiveras eller inaktiveras vid vissa stimuli. Genom att visa nätverket tusentals, eller till och med miljontals exempel på det du vill att den ska göra så kommer den anpassa sig och lära sig känna igen liknande saker.

Att veta vad en människa gör för någonting är lätt att förstå genom att titta på personen. Men exakt vad baserar du din tolkning på? Är det ansiktsuttrycket? Är det kroppsställningen? Eller är det kontexten som avgör? En djup och komplicerad blandning av allt detta är förmodligen svaret. Vi har fokuserat på att enbart tolka kroppsställningen som vi utvinnet genom ett neuralt nätverk som har lärt sig känna igen kroppsleder. Skelettet som utvinns kan ses i figuren under. Det utvunna skelettet matas sen vidare in i ett

annat neuralt nätverk som är specialiserad på att hitta mönster i kroppsställningen för att tolka vad personen gör.

Genom att utgå från ett tidigare stort dataset inom rörelsemönster, spelades ett nytt in med flera nya perspektiv för att undersöka hur prestandan på dessa vinklar kan förbättras. Genom att mata vårt neurala nätverk med filmer från dessa nya vinklar lyckades vi få nätverket att prestera nästan lika bra på den nya vinkeln som den gamla vinkeln. Detta utan att resultatet på den gamla vinkeln sjunker i någon större utsträckning. Resultatet ger en uppfattning om hur olika mängder data från nya vinklar påverkar prestandan vilket är en viktig förkunskap i utvecklingen av framtida applikationer inom området.



Figur 1: Bilder visar det skelett som utvinns av ett neuralt nätverk för att beskriva människokroppen. Detta skelett används sedan för att kategorisera olika rörelser. I detta fallet en person som kastar.