

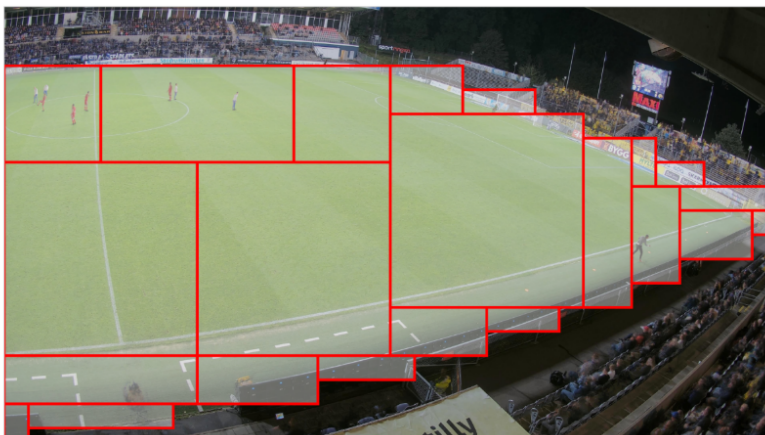
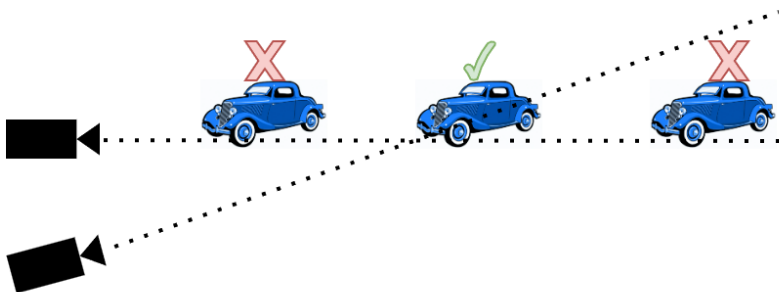
Populärvetenskaplig sammanfattning

De senaste åren har chatbotar, såsom ChatGPT och Le Chat, blivit välkända verktyg som fått symbolisera framstegen inom AI-forskningen. De har blivit kända för sin användarvänlighet och används både av forskare och allmänheten. Parallellt med chatbotar sker det emellertid även utveckling inom bland annat datorseende och maskininlärning, och där denna avhandling behandlar dessa ämnen.

Två huvudteman för avhandlingen är positionering samt rekonstruktion av objekt från kamerabilder. Genom att använda flera kameror, eller en kamera som tar många bilder från olika positioner kan man använda sig av en teknik som kallas triangulering. Vanliga kameror vi använder idag är så kallade projektionskameror, där samtliga punkter längs en linje i rummet alla kommer visas i samma punkt på bilden. Men det är endast punkten närmast kameran som kommer synas i bilden. Triangulering fungerar genom att vi vet att ett objekt (eller punkt) befinner sig längs en linje från en kamera, men inte var längs den linjen. En annan kamera kan då, från en annan vinkel, ha en egen linje som korsar den första. Då kan vi dra en slutsats om objektets faktiska position. Objektet och de två kamerorna skapar en triangel, därav namnet. En illustration av triangulering kan ses i övre halvan av Figur 0.2. Denna teknik kan vi människor med våra två ögon också använda i viss utsträckning även om större avstånd mellan kameror (eller ögon) tenderar att ge bättre resultat. Men vad händer om vi täcker för ett öga som en pirat? Efter lite invänjning klarar vi människor oss bra ändå. Denna avhandling har behandlat just det fallet då vi endast har en kamera, en så kallad monokulär uppsättning. Då saknas möjligheten till triangulering men genom addering av maskininlärning och andra lösningar kan detta delvis avhjälpas.

Artikel I behandlar positionering och rekonstruktion av bilar, sett från en annan bil. Detta är tilltänkt för exempelvis självkörande bilar där man vill kunna positionera samtliga trafikanter i en omgivning kring bilen. Förutom bilar finns även andra trafikanter, vilket artikel II adresserar, där fotgängare och cyklister positioneras tillsammans med bilarna. Detta kan användas för trafikövervakning där man automatiskt vill kunna detektera potentiellt farliga situationer eller olyckor.

Efterföljande artiklar behandlar monokulär positionering av fotbollsspelare på en fotbollsplan. Positioneringen gjordes som en del av att generera ett fågel-perspektiv (Eng: Bird's-Eye-View) där fågelperspektivet underlättar positioneringen. Metoden gick ut på att skapa ett rutnät över fotbollsplanen som man därefter för varje ruta, genom ett neuronnät, klassificerade om det fanns en fotbollsspelare i den eller inte. För att öka träffsäkerheten uppskattades även finjusteringar av positioner i samma neuronnät då detektion i en ruta endast kan bli så noggrann som rutnätets storlek. Det initiala arbetet gjordes i artikel III med en förbättring i artikel IV. Förbättringen gjordes med hjälp av lokal maxdetektion (Eng: Non-maximal suppression), som hindrar att flera detektioner görs på varandra. Ar-



Figur 0.2: Triangulering (övre halvan), sker då kamera 1, som är en projektkamera, kan se att en bil (eller en punkt av den) finns någonstans längs linjen, men inte var. Kamera 2 kan då med hjälp av sin information med en linje till bilen, lösa problemet med hjälp av triangulering. Indelning av en fotbollsplan i rutor (nedre halvan) behöver inte ha samma form eller vara kvadratisk utan är gjord för att minska beräkningarna och minnesanvändningen för den aktuella bilden. Bilden är ett exempel på resultaten från artikel V.

tikel V introducerar bland annat indelning av bilden som analyseras av nätverket (Eng: Tiling). Genom strukturen på nätverket kan vi garantera samma detektioner, fast över en liten del av fotbollsplanen i taget. Indelningen illustreras i nedre halvan av Figur 0.2 där varje ruta behandlas var för sig. Genom att behandla en del av bilden i taget kan minnesanvändningen hållas nere. Resultatet från varje del kan sedan slås ihop till detektioner för hela fotbollsplanen. Artikel VI avslutar avhandlingen genom att sätta ihop systemet från artikel III, IV och V samt att bland annat kvantisera data och vikter i neuronnet. Kvantisering innebär att man istället för att använda decimaltal, så kallade flyttal, endast använder heltal. Dessa är mycket enklare och snabbare att beräkna för nätverket och kräver mindre minne. En nackdel är att de kan försämra resultaten. När kvantiserade nätverk används kan algoritmerna användas på små, lokala enheter utan de kraftfulla grafikorterna som vanligtvis behövs.

Resultaten från artiklarna visar på möjligheten till att använda små och specialiserade nätverk för både positionering, där användningen kan ske utan stora serverhallar eller specialiserade stationära datorer.