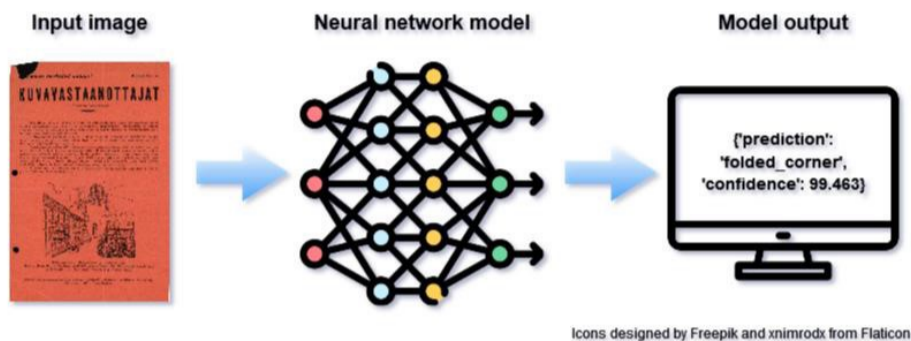


Igenkänning av vikta sidhörn

Bakgrund, fördelar och ändamål med komponenten

I processen för digitalisering av dokument kan det hända att dokument där ett hörn eller en kant på pappret är vikt eller sönderrivet överlämnas för skanning. Om det vikta eller sönderrivna sidhörnet påverkar läsbarheten av innehållet, är det bra om sådana sidor kan kännas igen och vid behov skannas på nytt. En maskinell igenkänning av vikta sidhörn bidrar därmed till att förbättra kvaliteten på digitalt material och minskar behovet av en manuell kvalitetskontroll.

Vad gör komponenten?



Komponenten behandlar de bildfiler som den fått som indata en efter en och tar fram en prognos för bildens innehåll på grundval av de parametrar som den har lärt sig under utbildningen. Om komponenten känner igen en vikt eller sönderriven bild med mer än 50 procents sannolikhet, klassificeras bilden som felaktig. Komponentens återställer den förväntade klassen till användaren.

Hur används komponenten?

Komponenten finns tillgänglig både som en del av Arkkiivi-gränssnittet (<http://www.arkkiivi.fi/>) och som en fristående app som tillhandahåller ett programmeringsgränssnitt (API) på DALAI-projektets GitHub-webbplats (<https://github.com/DALAI-project/CornerAPI>). Närmare anvisningar om användningen av komponenten inom olika miljöer finns på den webbplatsen.

Utbildning av komponenten

Bildfiler i .jpg-format har använts för att utbilda komponenten och deras storlek har harmoniserats till 224 x 224 pixlar. Det totala antalet dokument som använts i utbildningen uppgår till cirka 35 000 exemplar, av vilka cirka 5 000 är vikta eller sönderrivna dokument. Utbildningsmaterialet är utvalt så att komponenten kan känna igen vikta eller sönderrivna sidhörn i dokumenten, oberoende av om felen påverkar läsbarheten av dokumentet direkt (t.ex. behöver ett vikt sidhörn inte dölja textinnehållet).

Trots att man har försökt samla många olika exempel till utbildningsmaterialet, är materialmängden begränsad och det uppstår också fel i komponentens klassificering. Om ett dokuments hörn till exempel på grund av sin färg eller form liknar ett vikt sidhörn, kan detta leda till en felaktig klassificering.