INSTITUTIONEN FÖR MATEMATIK | LUNDS TEKNISKA HÖGSKOLA | PRESENTERAD 2018-11-23

EXAMENSARBETE Audio representation for environmental sound classification using convolutional neural networks
STUDENT Linus Lexfors, Malte Johansson
HANDLEDARE Kalle Åström (LTH)
EXAMINATOR Andreas Jakobsson (LTH)

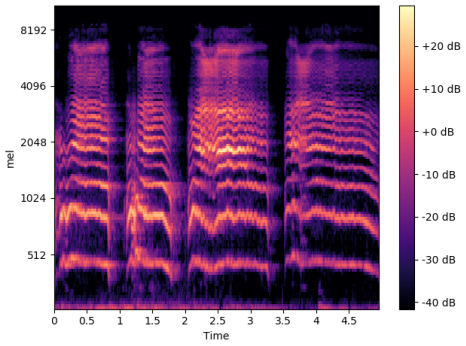
Datarepresentation för klassificering av ljud med neurala nätverk

POPULÄRVETENSKAPLIG SAMMANFATTNING **Linus Lexfors, Malte Johansson**

Neurala nätverk har visat starka resultat i bland annat bildklassificering. Men hur hanterar dessa tekniker ljudklassificering? I vårt arbete gör vi ljud till bilder, i form av så kallade spektrogram. Dessa används för att träna ett nätverk till att kunna skilja på 50 olika typer av ljud. Vår bästa modell har en precision på 74.7%.

Maskininlärning har praktiskt taget exploderat i popularitet under de senaste åren. Vi ser inte Terminator i framtiden utan tycker bara det är häftigt. Grundidén går ut på att låta ett system "träna", genom att visa det massor av exempel av sakerna man vill att det ska kunna känna igen. I början är systemet ute och cyklar helt, det gissar bara. Men eftersom vi vet det rätta svaret själva kan vi säga hur pass mycket fel det har. Har man sedan lyckats med sin design blir systemet bättre och bättre ju fler exempel det får bearbeta. Men man måste hålla tungan rätt i mun, det är viktigt att man inte utvärderar systemet med exemplena det har fått se under träningen. Istället håller man undan några exempel tills det är dags att se hur bra systemet hade funkat i nya situationer.

Designen av nätverket är inte det enda som påverkar resultatet, hur man representerar sin data spelar också stor roll. Den råa ljuddatan innehåller inte så mycket information i sig själv och är svår att rita upp som en bild på ett värdefullt sätt. För att få ut mer information kan man omvandla ljudet till ett så kallat spektrogram. Ett spektrogram innehåller information om vilka frekvenser som finns i ljudet och hur det förändrar sig över tiden. I vårt arbete testade vi olika sätt att ta fram dessa bilder. Den bästa modellen som fick arbeta med spektrogram fick en precision på 63.35%. Ett alternativt sätt att representera dessa ljudbilder är att använda sig utav en annan skala för frekvenser. Vår bästa modell använder en skala som kallas Mel-skalan och den trycker ihop höga frekvenser i bilden och sträcker ut de lägre. På detta sätt tar de låga frekvenserna upp mer plats i bilden och det blir lite lättare för modellen att hitta och känna igen mönster i dessa lägre band. När vi tränade med Mel-skalade spektrogram fick vi en precision på 74.7%. Vid en undersökning visade det sig att människor hade en precision på 81.3% på ljuddatan, läskigt nära va?



Mel skalat spektrogram utav ett gråtande barn.