

# Animarea modelelor 3D folosind rețele neuronale

Vavilov Andrei

June 16, 2021

## 1 Introducere

Inteligența artificială este un domeniu foarte versatil, a cărui aplicabilitate poate fi extinsă în foarte multe domenii, printre care și modelarea grafică, mai exact în modelarea și animarea 3D.

Următorul studiu își propune să prezinte o parte din modelele și tehnologiile care pot fi folosite pentru a anima un model 3D în timp real, pe baza unui fișier audio selectat de către utilizator.

## 2 Tehnologii propuse

Deoarece inteligența artificială este unul dintre cele mai intens cercetate domenii din prezent, există o varietate de tehnologii și instrumente pentru realizarea de proiecte.

În continuare voi prezenta și argumenta pe scurt alegerile făcute.

### 2.1 Limbajul de programare

O întrebare care a apărut odată cu creșterea în popularitate a domeniului inteligenței artificiale este “Ce limbaj de programare ar trebui să folosesc?”. Deși nu există “cel mai bun limbaj pentru inteligența artificială”, Python este de cele mai multe ori alegerea preferată atât a programatorilor cât și a oamenilor de știință și a statisticienilor datorită:

1. Varietatea de framework-uri și instrumente pentru Machine Learning, Data science și modelare/interpretare a datelor (TensorFlow, Scipy, Pytorch, Numpy, Pandas).
2. Numărul mare de resurse disponibile online

3. Sintaxa simpla
4. Fiind un limbaj *dynamically typed*, realizarea de aplicatii/modele este facila si rapida.
5. Fiind un limbaj *cross platform*, aplicatiile sunt automat suportate pe o varietate mare de platforme si sisteme de operare.

## 2.2 Tipul de *Neural Network* folosit

Inainte de a argumenta alegerea rețelei neuronale, trebuie introdus pe scurt conceptul de ***Deep Learning***.

*Deep Learning*, cunoscut si sub numele de *Deep Structured Learning*, face parte din familia invatarii automate si reprezinta o combinatie a algoritmilor de *Machine Learning* cu *Representation Learning* (tehnica de automatizare a interpretarii datelor). Principalul avantaj pe care *Deep Learning Neural Networks* il prezinta fata de algoritmii de *Machine Learning* este minimizarea intervenției factorului uman, prin automatizarea procesului de *feature extraction*, adica extragerea caracteristicilor distinctive ale datelor de intrare.

Din categoria *Deep Learning Neural Networks* trebuie amintite **ANN**, **CNN**, si **RNN**, principalele categorii de *Deep Learning Neural Networks*.

- ANN

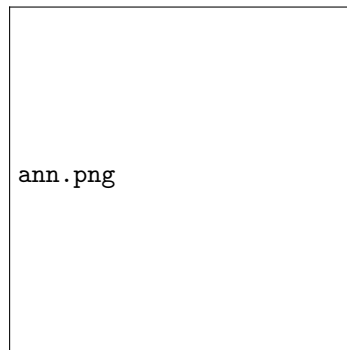


Figure 1: Ilustrare schematica a arhitecturii ANN

Retelele neuronale de tip ANN au ca si caracteristica principala prelucrearea unidirectionala a datelor (*forward feeding*) si este un tip de retea preferata in general in cazul problemelor din spectrul business-ului( predicții in legatura cu vanzarile/ viitoarele interese ale cumparatorilor, validare de date, risk management, etc.).

- CNN

Retelele neuronale de tip CNN sunt folosite in cadrul problemelor de clasificare a imaginilor, acestea prelucrand caracteristicile spatiale(pozitia pixelilor si relatiile dintre acestia).

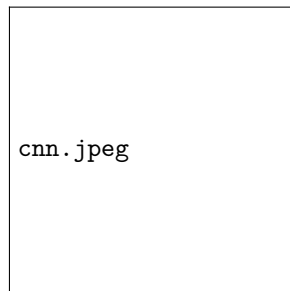


Figure 2: Ilustrare schematica a arhitecturii CNN

- RNN

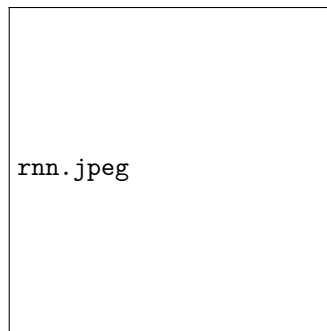


Figure 3: Ilustrare schematica a arhitecturii RNN

Retelele neuronale de tip RNN, desi au o structura relativ asemanatoare cu ANN, se deosebesc prin capacitatea de a prelucra datele in ambele directii (*recurrent feeding*), fiind astfel o optiune valida preferata in cazul problemelor de procesare text si audio. RNN transmite secvential informatia catre straturile de neuroni, determinand dependentele dintre componentele datelor de intrare. Aceasta proprietate distinctiva se numeste *Parameter sharing* si duce la scaderea numarului de neuroni necesari din retea si, implicit, la scaderea costurilor computationale.

### 3 Aprofundarea tehnologiilor propuse

Clasificarea instrumentelor muzicale nu este o problema noua, aceasta fiind abordata de cercetatori in invatare automata prin tehnici precum *SVM* (rata de eroare de 30% ), si clasificare prin *K-NN*. In ultimii ani insa cercetarea s-a indreptat spre retele neuronale deoarece acestea elimina nevoia unui factor uman care sa extraga trasaturile datelor de intrare. Apare insa problema alegerii tipului de retea neuronală. Articolul [DLFMIR] compara arhitectura *leNet CNN* cu arhitectura *LSTM* ( derivata din *RNN*), acestea fiind principalele candidate pentru rezolvarea problemei date.

Rezultatul experimentului a demonstrat ca, in urma antrenarii celor doua rețele neuronale cu o multime de date de intrare alcatuita din 200 de fisiere audio, cu lungime de aproximativ o secunda, arhitectura *leNet CNN* ajunge la o acuratete foarte apropiata de cea a *LSTM* mult mai rapid, in timp ce foloseste mult mai putine resurse, fiind deci o alegere mai buna pentru problema clasificarii de instrumente muzicale. Urmatoarea figura ilustreaza rezultatele obtinute in urma testului dupa un antrenament de 100 de epoci.

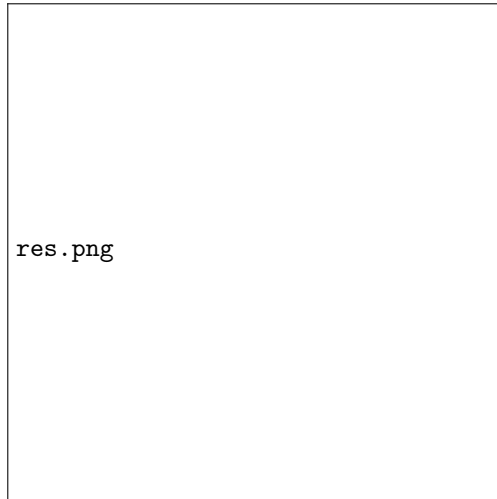


Figure 4: Rezultatele obtinute comparatiei dintre LSTM si leNet

Pentru a transforma fisierele audio in input valid vom folosi spectrograme *Mel*. Pentru a intelege cum functioneaza, este utila prezentarea conceptelor de *semnale audio* si *transformari Fourier*.

Un semnal reprezinta ilustrarea unui fenomen in timp. In cazul fisiereleor audio, semnalele indica variatia cantitatii de aer din timpul unei inregistrari. Aceasta variatie este in general determinata prin masurarea la intervale foarte mici (  $44.1\text{kHz} \approx 44100$  de inregistrari pe secunda).

*Transformarea Fourier reprezinta* o tehnica de transformare matematica care descompune o functie relativ la dependentele sale spatiale si temporale. Semnalul audio este compus din unde sonore care ilustreaza doar amplitudinea unei anumite inregistrari. *Transformarea Fourier* permite convertirea de la o ilustrare temporală la o ilustrare a frecventei, numita spectru.

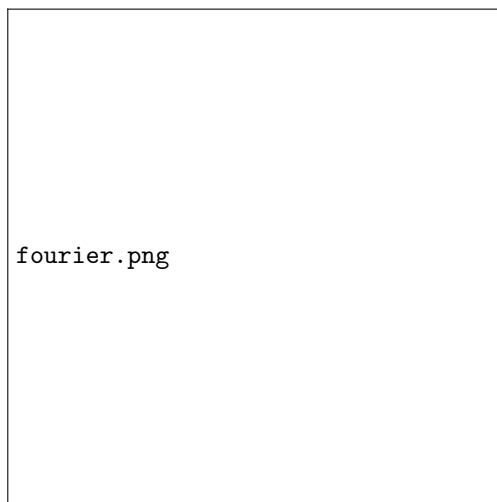


Figure 5: Procesul de transformare Fourier

O spectrograma este o reprezentare grafica a aplicarii algoritmului de transformare Fourier peste semnale nonperiodice. Spectogramele ilustreaza caracteristici ale sunetului precum nivelul de zgomot, amplitudinea, variatia frecventei, nivelul de decibeli etc.

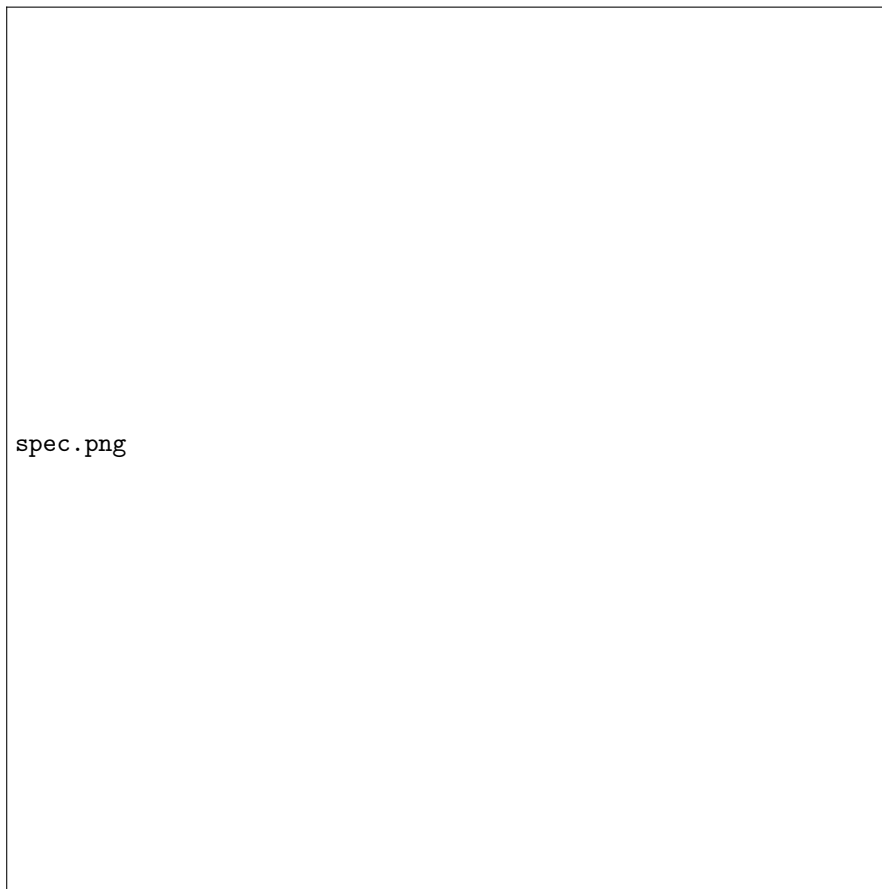


Figure 6: Spectograma audio

Scala *Mel* este o reprezentare a sunetelor perceptibile de catre om. Aceasta este folosita pentru normalizarea si eliminarea de caracteristici irelevante ale fisierelor audio.

Procesarea datelor de intrare pentru reseaua neuronală poate fi facilitata prin intermediul *Spectogra melor Mel*, transformand problema de recunoastere audio intr-o problema de recunastere a imaginilor.