Deep learning Házi feladat beszámoló

Készítette:
Benedek Zoltán, VZ9AS0
Biró Márton, Z7A244
Vizi Kristóf Levente, GN2VV4

Megajánlott jegyért.

Bevezető, célkitűzések, motiváció

A projekt célja neurális háló tanítása madarak fajtájának felismerésére a hanguk alapján. Szeretnénk egy olyan rendszert tanítani amely nagy pontossággal tudja felismerni a madárcsiripelés alapján a fajt ezzel segítve az adott régió biodiverzitásának feltérképezését.

A téma választásában nagy szerepet játszott továbbá, hogy mindhármunk számára ismeretlen terület volt a hangfájlokkal való szoftveres munka, így a neurális háló összeállításán kívül további kihívás is rejlett a feladatban.

Korábbi megoldások, előnyeik, hátrányaik

A projektünk kvázi a BirdCLEF 2024 Kaggle verseny feladványa volt. Mivel ezt a versenyt a Kaggle-ön már többedik éve teszik közzé, így rengeteg megoldás segítségünkre volt.

Ennek persze hatalmas előnye, hogy a korábbi megoldók zsákutcáit az esetek többségében ki tudtuk kerülni, de ez persze nem azt jelenti, hogy teljesen sikerült volna kiküszöbölnünk minden pluszmunkát.

A sok korábbi megoldás hátránya, hogy nehéz egy helyes utat választani, hiszen több különféle módszerrel is magas eredmények érhetők el. És persze nem szeretnénk egy az egyben lemásolni egyik megoldást sem. Ezek mellett fontos megjegyezni, hogy a különböző megoldások más-más, számukra jól működő részleteinek keverése közel sem biztos, hogy számunkra is javulást fog hozni, hiszen lehet, hogy az adott korábbi megoldás bizonyos része a készítőknek azért működött, mert teljesen más módszert választottak megoldásuk egy korábbi fázisában.

Az évek során a közzétett adathalmazokban nagyobb változások is történtek, így a hozzájuk tartozó legjobb megoldások is változtak. Ezek miatt mi elsősorban a

2024-es és 2023-as kiíráshoz érkező megoldásokból tudtunk szükség esetén ihletet meríteni.

Rendszerterv

A rendszer az alábbi komponensekből áll:

1. Adatok feltöltése Google Drive-ra
2. Vizualizáció
3. Adatelőkészítés: Az .ogg formátumú hangfelvételek, spektrogramokká alakítása, amelyeket a neurális háló bemeneteként használunk, majd a generált spektrogramok kiírása Google Drive mappákba.
4. Spektrogramok szűrése a Google Bird Vocalization Classifier használatával
5. Duplikált spektrogramok kiszűrése
6. Modell összerakása és tanítás:
    a. Transfer learning: előtanított EfficientnetV2B0 használata
    b. Általunk összerakott Dense és Dropout rétegek használata
7. Kiértékelés: Starfield 5-fold cross validation-nel

Adatbázis(ok)

Az adathalmaz madarak hangfelvételeiből áll .ogg fájlformátumban és 2 darab .csv file-ból melyek a hangfelvételekhez és a madárfajokhoz tartalmaznak információkat. A hangfelvételek között vannak címkézett és címkézetlen adatok is. A .csv fájlok a hangfelvételek helyszínét és a madárfajok megnevezését tartalmazzák.

A projektet szinte kizárólagosan a Google Colab és Google Drive alkalmazások használatával írtuk és futtattuk.

Architektúra, tanítás, nehézségek és megoldásuk

Az első nehézségünk a hangfájlok átalakítása volt feldolgozható formátumra. Itt több választási lehetőséget is kiértékeltünk: Mel spektrogram, MFCC, Wav2vec.
Az MFCC főleg emberi hang feldolgozására lett kialakítva, így a közép frekvenciákat hangsúlyozza. Mivel a madárcsiripelésben nagyon fontos szerepet játszanak a magas frekvenciák, így ezt a lehetőséget elvetettük.
Ugyancsak ígéretesnek tűnő, de viszonylag kevésbé bevett megoldás a Wav2vec, ez azonban jelentősen nagyobb számítási kapacitást igényel a másik két lehetőségnél.A Colab-os bottleneck miatt ezt sajnos nem engedhettük meg magunknak.
A 3 lehetőség közül végül a spekrogramos átalakítást választottuk, mivel széles körben bevett szokás a hasonló feladatokban. Erre így jelentősen több referenciát találtunk hasonló feladatok megoldásiban, illetve a Kaggle verseny legjobban teljesítő modelljei szinte kivétel nélkül ezt a megoldást használták.

A végleges spektrogram paraméterei:

- n_fft = 2048
- hop_length = 512
- n_mels = 128
- fmin = 40
- fmax = 16000

A madárhangok általában magasabb frekvenciájúak, ezért van az, hogy az alacsony frekvenciákat kiszűrjük, a magasakat pedig meghagyjuk.

Második nehézségünk az adathalmaz hatalmas méretéből adódott. 182 osztály, a legnagyobbakban akár 500 hangfelvétel, melyekből a legnagyobbak akár 4 perc hosszúak is lehetnek. Ez rengeteg adat. A spektrogramok generálása során rá kellett jönnünk, hogy tarthatatlanul sok időbe telik az első ötletünk - a hangfájlok 5 másodperces részekre bontása, 2,5 másodperces átlapolással. Szerencsére a BirdCLEF 2024 versenyzői rájöttek, hogy a hangfájlokban túl sok a zaj (pl. a csicsergés nincs is benne a nagy részében), ha a teljes felvételeket fel akarjuk használni, de minden felvétel első 5 másodperce meglepően nagy eséllyel tartalmaz tisztán hallható madárfüttyöt. Így minden felvételnek csupán az első 5 másodpercét használjuk fel.

Az adathalmaz 5 fold-ra való osztásakor az első fold jelentősen jobb eredményeket produkált. Ennek részletesen utánaolvasva kiderült, hogy a maradék felvételek sajnos továbbra is viszonylag sok zajos adatot tartalmaznak (az első fold kevesebbet, mint a többi), így a Google Bird Vocalization Classifier segítségével kiszűrtük azon spektrogramokat, melyek label-je nem hasonlított a Google Classifier által adott eredményre.
A szűrés során volt osztály, amiben alapból is nagyon kevés adat szerepelt, és egyik sem volt tiszta, ennek következtében pedig teljesen megsemmisült. Ezen osztály példányait szűrés előtti állapotukba visszahelyeztük.

A tanítás előtt még átvizsgáltuk az adatokat, hogy ne szerepeljenek közöttük duplikáltak.

A megmaradt adatok közül azon osztályokat, melyek számossága kisebb, mint az osztályok átlagos sámossága, augmentáljuk az alábbi technikákkal, 0.5 eséllyel:

- Time masking
- Frequency masking
- Coarse Dropout (négyszög tartományok eldobása)

A neurális háló architektúrája a következő: A legsikeresebb Kaggle megoldásokat tanulmányozva transfer learning-et használtunk, aminek alapjául az előtanított EfficientnetV2B0 verzióját vettük.
E mögé csatolunk egy 128-as Dense réteget ReLU aktivációval, és egy Dropout

réteget, hogy megakadályozzuk a túltanulást. A kettő között pedig végrehajtunk egy Batch normalizálást.
A modellt először az Efficientnet fagyasztásával tanítjuk, majd finomhangoljuk úgy, hogy visszavonjuk a fagyasztást.

Természetesen checkpointingot és early stoppingot is megvalósítottunk.

Eredmények és ezek kiértékelése

A háló pontosságának meghatározására a többosztályos osztályozás miatt kereszt-entrópiát használunk metrikaként. Ezen kívül a loss és accuracy metrikákat vettük figyelembe, mikor a modell régebbi és újabb változatait hasonlítottuk össze.

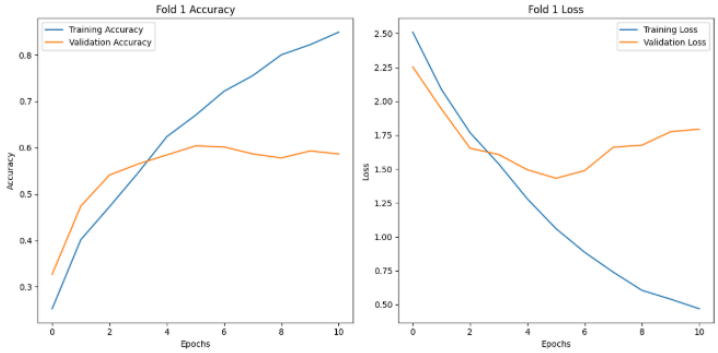
A hálót a Starfield K-fold Cross Validation módszerrel teszteljük, 5 fold-dal, hogy minden rendelkezésünkre álló adatot fel tudjunk használni.

DEMO

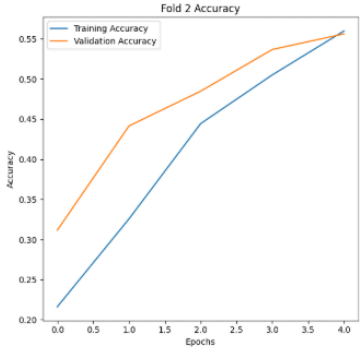
A képeken a legutolsó sikeres tanításunk futtatásának eredményei szerepelnek.
Az első képen az accuracy és loss metrikák láthatók. Összehasonlításképpen második képnek megmutatjuk, a projekt elején honnan indultunk.

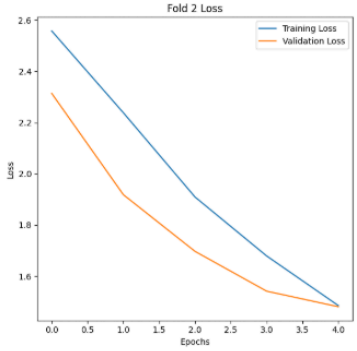
```
Cross-Validation Results:
Mean Validation Loss: 2.2019168376922607
Mean Validation Accuracy: 0.32676000595092775
Fold 1 - Validation Loss: 1.6671907901763916, Validation Accuracy: 0.5
Fold 2 - Validation Loss: 2.274763345718384, Validation Accuracy: 0.3181818127632141
Fold 3 - Validation Loss: 2.3957104682922363, Validation Accuracy: 0.2321041226387024
Fold 4 - Validation Loss: 2.3181521892547607, Validation Accuracy: 0.3080260157585144
Fold 5 - Validation Loss: 2.3537673950195312, Validation Accuracy: 0.27548807859420776
```

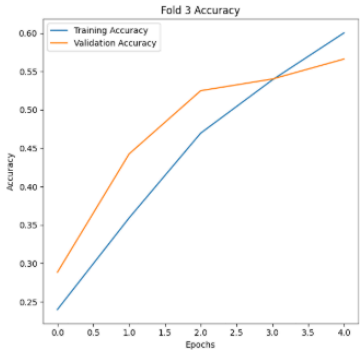
Alább pedig az egyes fold-ok pontosságának növekedése látható összehasonlítva a tanító és a validációs halmazokra.

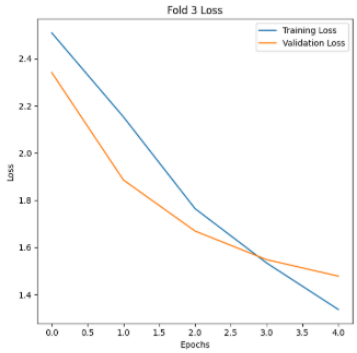

Fold 1 Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epochs

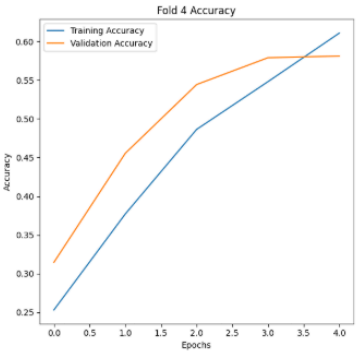
Fold 1 Loss
Training Loss
Validation Loss
Loss
Epochs

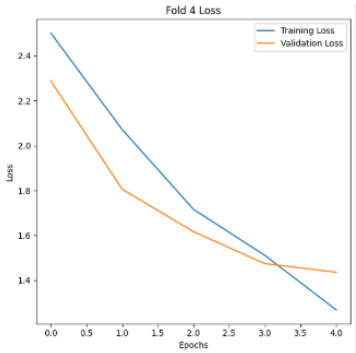

Fold 2 Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epochs


Fold 2 Loss
Training Loss
Validation Loss
Loss
Epochs

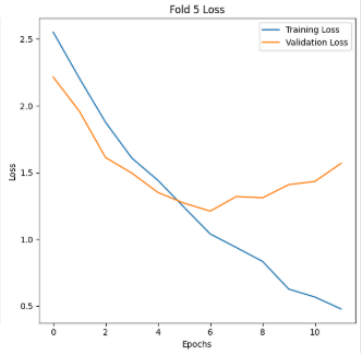

Fold 3 Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epochs


Fold 3 Loss
Training Loss
Validation Loss
Loss
Epochs


Fold 4 Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epochs


Fold 4 Loss
Training Loss
Validation Loss
Loss
Epochs


Fold 5 Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epochs


Fold 5 Loss
Training Loss
Validation Loss
Loss
Epochs

Összefoglaló

A projekt összességében egy kellően nagy, de természetesen nem lehetetlen kihívás volt számunkra. Élveztük is, de voltak stresszesebb pillanatai is.

A konkrét neurális hálók témakörén kívül rengeteg tudást, és főleg gyakorlati tapasztalatot összeszedtünk a nagy mennyiségű adatokon való dolgozásról, az audiófájlokkal való műveletvégzésről és a hivatalos dokumentációk segítségével történő munkáról is.

A feladat tagadhatatlan kihívása volt a Google Colab környezet, ami előfizetés nélkül ominózusabb projektek végzésére a korlátozott memóriaterülete és számítási kapacitása miatt igencsak nem alkalmas. Ez az akadály jött velünk szembe talán legtöbbször.

A projektben való haladást a diverz Kaggle megoldásokon kívül LLM is segítette. Ugyan a ChatGPT a ‘write this for me’ kérdésekre egyre jobb válaszokat ad, volt, hogy több problémát okozott, mint amennyit segített. Leginkább cikkek megtalálásában, implementációs ötletek adásában, könyvtárak *közel* helyes használatának megmutatásában, illetve nehezen észrevehető hibák megtalálásában volt segítségünkre.

És persze mindezek mellett egymás munkamorálját is kitapasztalhattuk a csapatban dolgozással.
A projekten való munkamegosztás a következőképpen zajlott:

- Benedek Zoltán: Eredeti augmentáció megírása, a modell összerakása, adatvizualizáció, spektrogramok szűrése a Google osztályozóval. Ő írta meg a projekt alappilléreit.
- Biró Márton: Adatvizualizáció, Kaggle-Drive-Colab környezetek összekötése, adatfeldolgozás, a projekt alapjának finomhangolása, részegységek integrálása, dokumentálás. A csapat koordinátora.
- Vizi Kristóf Levente: Adatfeldolgozás, és -generálás, hibajavítás, statisztika készítés, dokumentálás. A csapat processzora, aki éjjeli bagoly módjára futtatta egymás után a notebookokat.

Jövőbeli terveink: A modell bővítése, áthelyezése egy erőteljesebb környezetbe. A kész modell valós időben/környezetben való kipróbálása.