

# Retinaképek feldolgozása mély tanulási algoritmusokkal

Önálló projekt I.

Vas Bernadett

Az önálló projektem első félévében retinafelvételeket tartalmazó adathalmazon vizsgáltam különböző mély neurális hálókat viselkedését. A modellek feladata a diabetikus retinopátia detektálása volt, vagyis bináris klasszifikáció elvégzése az adathalmazon. A látást veszélyeztető elváltozások a cukorbetegség egyik szövődménye, korai diagnosztizálása rendkívül fontos a megfelelő kezelések elkezdése érdekében, amivel megelőzhető a súlyosabb látásromlás. Vizsgálataim középpontjában a különböző hálótípusok megértése, és retinaképek osztályozására való alkalmazhatóságuk állt. Az alapszakos szakdolgozatomban már megismertem két architektúrát, a ResNet-101-et és NFNet-et, ezúttal pedig más konvolúciós hálókat megértésére fókuszáltam.

A projekthez egy, diabetikus retinopátia felvételek klasszifikálására kiírt Kaggle verseny EyePACS adathalmazát használtam fel. A benne található képek színesek, és összességében zajos adathalmaznak mondható, mivel a felvételek többféle kamerából származnak, számos rossz minőségű képet és annotálást tartalmaz. Az adathalmaz osztályainak kiegyensúlyozása céljából az egészséges és beteg képekből is ugyanannyit válogattam be a tanítóhalmazba.

A tanított modellek között szerepeltek már régebb óta elterjedt konvolúciós hálókat, mint a VGG-16 és ResNet-101, valamint újabbnak számító, és az ImageNet adathalmazon state-of-the-art eredményt elérő architektúrák is, mint az NFNet, EfficientNet-B0, EfficientNet-B6 és EfficientNetV2-S. Mindegyik modell esetében tanításkor fine-tuning technikát alkalmaztam, azaz ImageNet-en tanított súlyokat használtam súlyinicializálásként minden rétegben, kivéve az utolsó klasszifikálóban, ahol véletlen értékadással voltak inicializálva a súlyok.

A VGG-16 modellnél [2] használták először azt a megfigyelést, hogy több kisebb kernelméretű konvolúciós réteg egymás utáni alkalmazása ugyanazt az eredményt tudja elérni mint egy-egy nagyobb kernelméretű konvolúció. A több réteg megnövekedett mélységet eredményez, ami komplexebb jellemzők megtanulását teszi lehetővé.

A ResNet-101 modell reziduális blokkokat használ, ami a gradiensek jobb áramlását biztosítja, és segít az eltűnő gradiensek problémájának megoldásában. Az NFNet háló a batch normalizáló réteg helyett más technikákat alkalmaz a gradiensek skálázására és a rétegek bemeneti eloszlásainak kordában tartására, azonban ezt jelentős paraméterszám növekedés mellett sikerül elérnie.

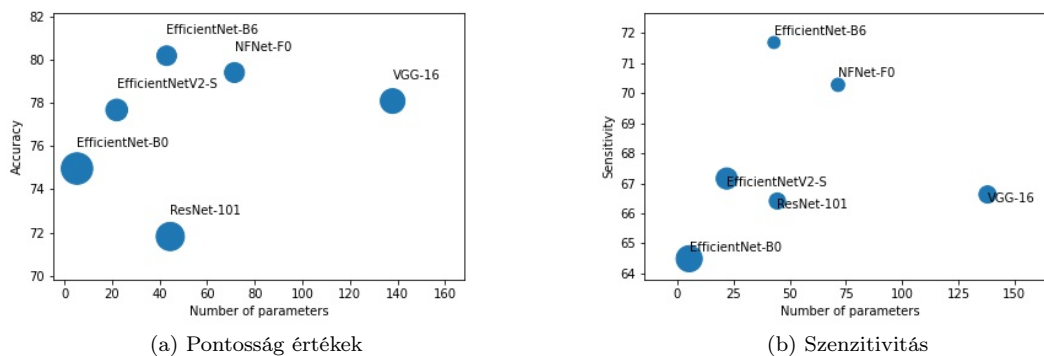
Az EfficientNet modellcsalád [3] architektúrájának hátterében az a feltevés áll, hogy óvatosan kell egy konvolúciós háló méreteit felnövelni. A felskálázással nagyobb variánsokat lehet létrehozni, és így egy egész modellcsaládot elkészíteni. Az irodalomban megtalálható architektúra növelés általában csak külön a mélységre, szélességre, illetve a képméretre szorítkozik. Ezek a paraméterek összefüggnek, ugyanis például nagyobb képmérethez érdemesebb mélyebb architektúrát építeni, hiszen rögzített kernelméret mellett több pixel kerül a kernel által fedett részbe, ami azt jelenti hogy több réteg kell a komplex jellemzők megtanulásához. Ugyanakkor a modellt szélesíteni kell, azaz a csatornaszámot növelni, ha minél finomabb jellemzők megtanulására fókuszálunk. Az EfficientNet családban összehangoltan történik a szélesség, mélység és képméret növelése is a nagyobb modellek megalkotásához, amik a B0, B1, B2, B3, B4, B5, B6, B7 verziók, növekvő sorrendben. A háló fő építőeleme a *mobile inverted bottleneck MBConv*, ami a MobileNetV2 modelltől [1] származik. Az invertált reziduális blokk abban különbözik a rendes reziduális bloktól, hogy amíg utóbbinál a blokk bemenete és kimenete sokcsatornás, a reziduális ágon pedig csökken a csatornaszám, addig az invertált esetében ez pont fordítva történik. Ezenkívül a blokk reziduális ágán *depthwise* konvolúció helyezkedik el. Ebben a konvolúcióban külön kezeljük a csatornákat, és mindre külön filtert alkalmazunk.

Az EfficientNetV2 [4] architektúra az EfficientNet továbbfejlesztése. Ennél a modellcsaládnál is az alapháló a paraméterhatékonyságra és pontosság növelésére lett optimalizálva, de a skálázása eltér az EfficientNet-étől. A konvolúciós hálókat architektúrái komponensekre vannak osztva, és az egyes komponensek nem ugyanolyan mértékben járulnak hozzá a tanulási sebességhez a gradiensek számolása miatt. A modellben ezért a nagyobb hálókat létrehozásakor az utolsó komponensekben növekszik fokozatosan a rétegszám. Ez a nézőpont hasonlít a ResNet skálázásához, de ott a komplexebb jellemzők megtanulása miatt történik a későbbi komponensek növelése. Ezzel szemben az NFNet minden komponenseben azonosan növeli a rétegszámot a nagyobb variánsok létrehozásához, hasonlóan az EfficientNet-hez, hogy teret engedjen a korai rétegeknek az egyszerűbb jellemzők kinyeréséhez.

A modellek teljesítményének mérésére több metrikát is használtam, hogy árnyaltabb képet kapjak az elért eredmények viszonyáról. Ahhoz, hogy a modellek eredményeit ne torzítsák el a tanítás során megjelenő randomizálások, tízszer tanítottam minden modellt, és a teszhalmazon az adott metrikában elért eredmények átlagát és szórását is tekintettem. A szórás meghatározása egy mutatót is adott arra, hogy mennyire tud stabilan osztályozni egy modell. A modellek összehasonlításához használt metrikák a pontosság és szenzitivitás. Utóbbi az orvosi diagnosztizálás szempontjából fontos mérőszám, azt méri, hogy az összes pozitív esetből hányat prediktált helyesen pozitívnak a háló. Mivel szeretnénk mindenképpen elkerülni, hogy egy beteg emberre mondjuk hogy egészséges, így fontos, hogy magas szenzitivitása legyen a modellnek. A kapott eredmények az 1. ábrán láthatóak. A kék körök középpontjai jelentik az elért átlagot, a kör sugara pedig a szórást mutatja. Az EfficientNet-B6 modellnek van a legmagasabb pontossága nem túl nagy paraméterszám mellett, ezekívül stabilnak is mondható a teljesítmény, mert relatívan nem túl nagy a szórása sem. Ezek a tulajdonságok fontosak lehetnek abban az esetben, ha orvosi eszközökbe, esetleg mobiltelefonba való beépítés jön szóba, és hatékony diagnosztizáló algoritmusra van szükségünk limitált tárhely mellett.

A szenzitivitás ábráján egyértelműen az EfficientNet-B6 és az NFNet teljesített jobban, és ezek rendelkeznek relatíve a legkisebb szórással is. Mindkét architektúrában található SE blokk, ami a csatornák közötti kapcsolatok megtanulásával növeli a reprezentációs képességet. Ezt úgy teszi meg, hogy a bemenet csatornáit súlyozza, ezeket a súlyokat pedig sűrű rétegekkel és nemlinearitásokkal állítja elő. Mivel ez a két háló ismert fel jobban és biztosabban a beteg osztályt, felvetődhet, hogy a megfelelő mélység mellett ebben szerepe lehetett a beépített SE blokkoknak is.

A következő félévben érdemes lenne áttérni bináris klasszifikációról az adathalmazban is megjelenő 5 osztályú klasszifikálásra, ami már nem csak a betegség detektálását jelenti, hanem annak a súlyosságának meghatározását is. Mivel a többosztályú címkék egy romló folyamatot írnak le, ez regressziós feladatként is felfogható. Az új feladat által azt is láthatnánk, mely súlyossági osztály megtanulása a legnehezebb, illetve melyek megtanulása segíthet a nehéz képek felismerésében is. Ezenkívül, mivel az EyePACS a legnagyobb publikusan elérhető adathalmaz a feladatra, felvetődik a kérdés, hogy vajon elég diverz-e ahhoz, hogy a rajta tanított modellek jól, illetve jobban teljesítsenek más retinafelvételes adathalmazon. Hiszen egy jól működő modellnek nem csak egy kórház adathalmazán kell működnie, hanem másokén is, hogy használható lehessen.



1. ábra. Eredmények

## Hivatkozások

- [1] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. MobileNetV2: Inverted Residuals and Linear Bottlenecks, 2019.
- [2] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition, 2015.
- [3] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, 2020.
- [4] Mingxing Tan and Quoc V. Le. EfficientNetV2: Smaller Models and Faster Training, 2021.