



**Karsa Róbert**

## TÚZESETEK FELDERÍTÉSÉNEK ÚJ LEHETŐSÉGEI

### Absztrakt

A cikkben a tűzesetek és műszaki mentések vizuális információk felhasználásával történő felderítésének új gépi tanulás alapú módszereit vizsgálom meg. A célom az volt, hogy mesterséges intelligencia módszereket felhasználva utcai kameraképek feldolgozásával kinyerjek olyan információkat a képekből, amelyre korábban csak emberek voltak képesek. Ezeket az információkat pedig tűzvédelmi célokra használtam fel, azaz a káresetek automatikus felismerésére. A kutatásom során internetről letöltött képek segítségével tanítottam egy modellt a tűzesetek, közlekedési balesetek felismerésére. A modellt Python nyelven a Pytorch keretrendszer segítségével valósítottam meg. Az általam létrehozott modell elég robusztus lett ahhoz, hogy automatikusan nagy magabiztossággal felismerje a tűzoltói beavatkozást igénylő eseményeket.

**Kulcsszavak:** gépi tanulás, kamerák, tűzeset, python, pytorch

## NEW POSSIBILITIES OF FIRE DETECTION

### Abstract

In this paper I explore new machine learning-based methods for detecting fires and technical rescues using visual information. My goal was to use artificial intelligence methods to extract information from street camera images that only humans could do before. This information was then used for fire protection purposes, i.e. automatic detection of damage. In my research, I used images downloaded from the internet to train a model to recognise fires and traffic accidents. I implemented the model in Python using the Pytorch framework. The model I created has become robust enough to automatically detect events with high confidence that require firefighter intervention.

**Keywords:** machine learning, cameras, fire, python, pytorch



## 1. BEVEZETÉS

### 1.1. Tűztorony

Az elmúlt évszázadokban a tűz elni védekezés ikonikus szimbóluma volt a tűztorony, amely a toronyőrök segítségével nemcsak a tüzek észlelésére szolgált, hanem biztonsági funkciói is voltak. A környezet megfigyelése, ellenőrzés alá vonása nem újkeletű dolog, csak a módszerek változtak.

### 1.2. Kamerák mindenhol

Napjainkban az élet minden területén feltűntek különböző kamerák, hogy megfigyeljenek, rögzítsenek eseményeket, folyamatokat. Ezek a modern kor tűztoronyai. A világot átölelő hálózat, az Internet segítségével ezek a képi információk közzétehetőek, így követhetőek és elemezhetővé váltak. A kamerák telepítésének egyik fő célja a biztonság elősegítése, megteremtése, fenntartása. A biztonság dimenziói komplex rendszert alkotnak, magukban foglalva többek között a társadalmi, gazdasági, ökológiai biztonságot, továbbá a tűzvédelmet, ha nem is főkategóriaként. Jelen dolgozatban a kamerák tűzvédelmi, tűzoltói felhasználásával foglalkozom, de mint azt látni fogjuk, ennél jóval nagyobb tér nyílik meg előttünk a lehetséges alkalmazások tekintetében.

### 1.3. Dokumentált világ, avagy hogyan tudom ezt felhasználni?

A minket körülvevő világ még sohasem volt ilyen jól dokumentálva, mint napjainkban. Kamerák figyelik a városok szinte minden részét, a gyárak, üzemek, üzletek belső tereit, a parkokat és még hosszasan sorolhatnám a listát. Ezekből a megfigyelésekből rengeteg információ keletkezik és ezeknek a többségét nem dolgozza fel senki, mert számítási kapacitás, ismeret vagy éppen a humán erőforrás hiányzik. Azt könnyen beláthatjuk, hogy emberként ezeket az információkat képesek vagyunk ugyan feldolgozni, de az ehhez szükséges emberi erőforrások nem állnak rendelkezésre vagy éppen gazdasági szempontból



rendkívül költséges lenne a fenntartásuk (több száz főt alkalmazni a kameraképek megfigyelésére). Ez korunk egyik fő kihívása, a „BigData” probléma. [1]

Az adatok értő feldolgozása számítógépek segítségével ma már nem „science fiction”, hanem egészen elérhető közelségbe került. Szinte mindenkinek ott lapul a zsebében az okos telefonnak nevezett eszköz, amely számítási kapacitását tekintve messze felülmúlja a néhány évvel ezelőtti személyi számítógépeket. A megfelelő számítási kapacitáson túl pedig rendelkezésre áll egy sor olyan módszer, amelyek segítségével az adatokat értő módon tudjuk elemezni. A mesterséges intelligencia egyik részterületének tekinthető gépi tanulás segítségével az adatok értő módon történő feldolgozása vált lehetővé és ezt szeretném bemutatni a következőkben.

#### **1.4. A modern kor tűztornyai**

Felmerül a kérdés, hogy amikor egy kamera éppen rögzít egy tüzesetet, vagy közlekedési balesetet, akkor ezt a tényt egy gép fel tudná-e ismerni, és amennyiben igen, akkor ezt milyen gyorsan és magabiztosan tudná megtenni. Ezentúl persze még számos kérdés is felmerülhet, de ezt az információt el kell tudni juttatni a megfelelő helyre. Ez utóbbi probléma már nem jelent igazi kihívást, csupán egy informatikai feladat, amit 10 évvel ezelőtt is képesek voltunk kezelni. A dolgozatban az első kérdésre fogok választ keresni, hiszen ez a probléma 10 évvel ezelőtt még a „science fiction” kategóriába esett, azonban ma már kezelhető feladat. Az angol nyelvű szakirodalomban „state-of-the-art” ként titulált, magyarul a legkorszerűbb technikának nevezném, azokat az eszközöket, amelyekkel ezek a feladatok megoldhatóak.

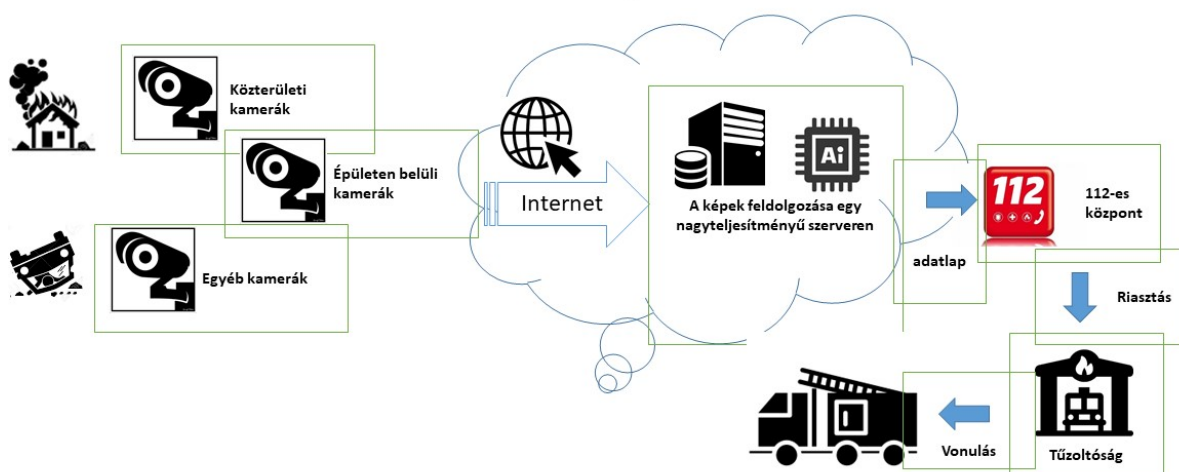
## **2. ELMÉLET**

### **2.1. Rendszertervezés**

Ebben a részben azokat a megfontolásokat vesszük sorra, amelyek mentén a rendszert tervezzük. A fentiekben megnevezett célkitűzés, azaz tüzesetek, műszaki mentések felismerése képi információkból, matematikai értelemben egy osztályozási feladat, amelyet jelenleg a leghatékonyabban úgynevezett mély neurális hálózattal lehet megvalósítani. A mély



neurális hálózatok közül ki kell választani egy architektúrát, amelyből egy modellt tudunk létrehozni és később alkalmazni. Az architektúra itt egy konkrét típust jelöl a neurális hálózatok nagy halmazából, amit külön fejezetben fogok bemutatni. Az architektúra kiválasztásánál sok szempontot kell figyelembe venni, itt most csak a futtatásához szükséges környezetet vizsgálom meg. El kell dönteni, hogy a modell hol fusson. Ez lehet a végpontokon, azaz a kamerákba integrálva, de lehet egy adatközpontban (felhőben), esetleg egy általunk közvetlenül felügyelt lokális szerveren. Egy okos telefonon futó szoftver már képes megkülönböztetni és osztályozni a képeket, pl. virágok, személyek, járművek, stb... egy fotón. A jelenleg kapható kamerák között egyre több úgynevezett okoskamera kapható, amely képes a kamerakép bizonyos feldolgozására, osztályozási feladatokra is. Ezek az okos kamerák azonban általában zárt szoftverrel rendelkeznek, azaz nem tudjuk saját igényeink szerint alakítani őket. Az egyik lehetséges megoldás a kamerák mellé egy célszámítógépet tervezni (edge computing), amely képes futtatni a modellt, majd a predikciót továbbítja az adatközpont felé. Ezzel a megoldással az a fő probléma, hogy jelenleg nagyon kevés olyan kamera van forgalomban, amelyet rugalmasan tudnánk programozni, illetve a már telepített kamerák többsége nem ilyen. További nehézség lehet egy modell frissítés (szoftver frissítés hasonlóan a telefonos applikációkhoz), amikor a jobb tulajdonságú modellt szeretnénk bevezetni. Ezt nehéz kivitelezni, hiszen minden egyes eszközön külön kell a frissítést elvégezni.



1. ábra: Tűztorony modell (saját szerkesztés)

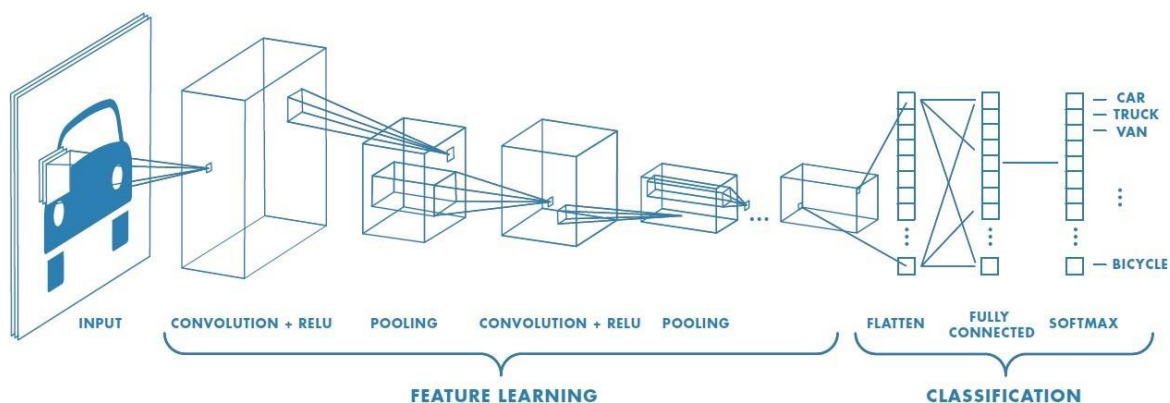


Másik lehetséges megoldás (1. ábra), hogy a kamerák képeit vigyük be egy nagy teljesítményű központi számítógépbe, ez lehet a felhőben is, de célszerűbb biztonsági okokból egy általunk felügyelt lokális szervert alkalmazni, majd a képeket itt feldolgozni. A dolgozatban is ezt a módozatot fogom bemutatni, hiszen ez tűnik a jelenlegi helyzetben a legjobban kivitelezhetőnek és legrugalmasabb megoldásnak.

El kell még döntenünk, hogy hány osztályt szeretnénk egymástól megkülönböztetni. A mély neurális hálózatok képesek akár 1000 db különálló osztály felismerésére is, de ilyen esetben - mint azt a következőkben látni fogjuk - nagy mennyiségű úgynevezett tanító adat szükséges, amelyet rendkívül munkaigényes összeválogatni. Ezért úgy döntöttem, hogy a dolgozatban csupán három osztályt fogok megnevezni, ezek a tűzeset, közúti baleset és utcakép lesznek, azaz városias környezetben működő kamerák képesek legyenek megkülönböztetni egy átlagos utcaképet egy tűzesettől.

## 2.2. Az architektúra kiválasztása

A számítógépes képfeldolgozási feladatokhoz fejlesztették ki az úgynevezett konvolúciós neurális hálózatokat, amelyek azon az ötleten alapulnak, hogy a hálózat különböző rétegei a kép egyes vonásait tanulják meg, és ezeket kombinálva annak egyre bonyolultabb részleteit képesek felismerni. [2]

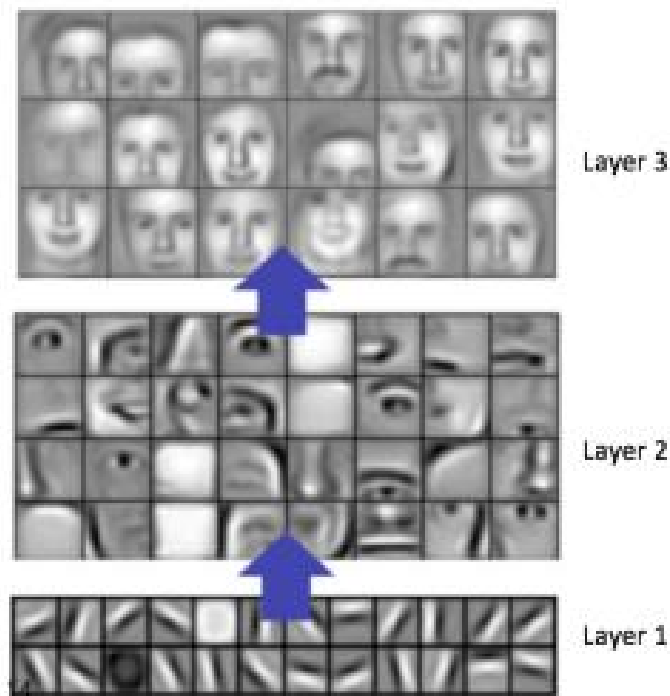


2. ábra: Konvolúciós neurális hálózat

Ez a folyamat látható az 2. ábrán. A hálózat első része a kép tulajdonságait (vonásait) tanulja meg, majd a hálózat második része végzi az osztályozási feladatot. A konvolúciós hálózat teljes bemutatására nincsen lehetőség jelen dolgozatban, azonban intuitív módon szemléltetve

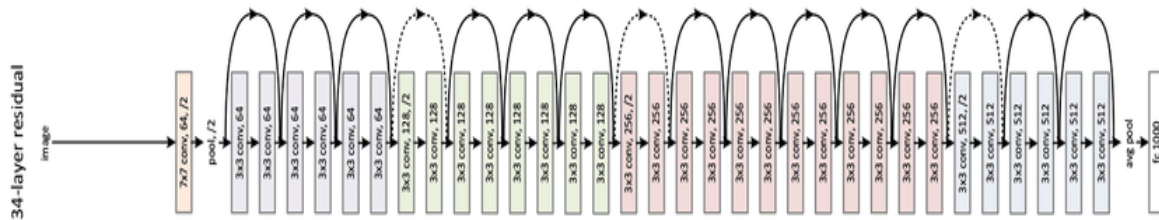


megérthető annak működése. A hálózat rétegekből áll, és az adatoknak a rétegeken kell keresztül menni, ami egyben az adatok feldolgozását is jelenti. A hálózat első rétege a bemeneti réteg, itt adjuk át az információt a hálózatnak, ami esetünkben egy kép. Az első rétegek még csak egyenes, ferde vonalakat ismernek fel, majd a következő réteg az előzőekből összeálló magasabb szintű jellemzőket képes kinyerni. A konvolúció az információk speciális kivonása a képből. Ezt a folyamatot figyelhetjük meg a 3. ábrán. Általánosságban elmondható, hogy minél több konvolúciós lépést alkalmazunk, annál összetettebb funkciókat képes megtanulni a hálózat. Az egyes rétegek által kinyert funkciókat a későbbiekben a hálózat képes kombinálni és felismerni bonyolult formákat a kép tetszőleges részén. A hálózat egyes rétegei a képnek egy sajátos reprezentációját tárolják és tanulják meg. [2]



3. ábra: Konvolúciós neurális hálózat rétegeinek vizualizációja

Az architektúra kiválasztásánál a konvolúciós neurális hálózatok közül a ResNet architektúrára esett a választásom.[3] A hálózat felépítését a 4. ábrán láthatjuk.



K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016

#### 4. ábra: ResNet-34

A ResNet architektúra 2015-ben készült el egy képosztályozó versenyre (ILSVRC 2015), ahol 1,2 millió képből álló tanítóhalmazt adtak közre, amelyek 1000 képosztályt fedtek le. Az elkészült modell teljesítményét 100 ezer, korábban nem „látott” képen mérték le. Érdekesség, hogy a modell pontossága meghaladta az ember által elért értéket is. [4]

Ez az architektúra megfelelően skálázható, rugalmas és robusztus az általam meghatározott feladathoz, továbbá elérhetőek az előre kiképzett változatai is. Az általam választott modell nem a legkorszerűbb, azonban a rendelkezésemre álló számítógéppel már kezelhető.

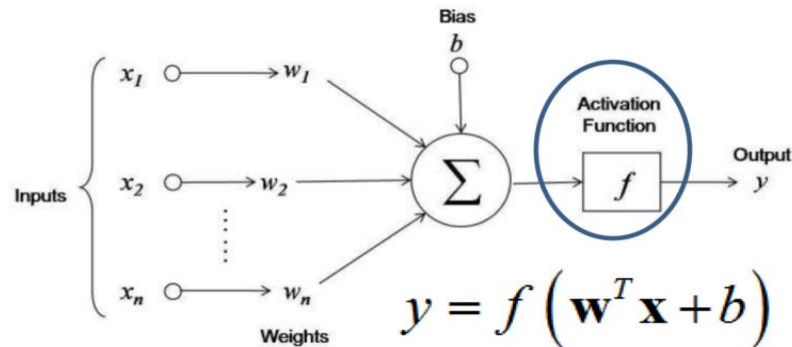
### 2.3. Egy neuron működésének bemutatása

A ResNet architektúrának több változata is elérhető, ezek a rétegek számában különböznek egymástól. A ResNet-34 34 rétegből épül fel, a ResNet-50 pedig 50 rétegből, és vannak még ennél nagyobb rétegszámú architektúrák is. A nagyobb rétegszám több képzendő paramétert jelent, így nagyobb erőforrás szükséges a modell tanításához. A ResNet-50 architektúrának több mint 23 millió képezhető paramétere van. [5]

Egy modell képzése során a modell paramétereit (számokat) kell beállítani, úgy, hogy az adott bemeneti értékekre az elvárt kimeneti értéket kapjuk. A feladat matematikai értelemben egy függvény illesztési feladat. Ennek a megértéséhez a következőkben megvizsgáljuk a legegyszerűbb alkotó elemet, a neuront. Az 5. ábrán a Perceptron modell látható, amelyet az emberi idegrendszer ihletett. A neuronhoz tartozik  $n$  darab bemenet ( $X_1$ - $X_n$ ), ezeken keresztül kap adatokat (ingereket) más neuronoktól vagy esetünkben bemeneti adatokat. A



bemenetekhez rendre  $w_1-w_n$  n darab paraméter tartozik, amelyek az egyes bemeneti kapcsolat erősségét modellezik (ezeket kell majd a későbbiekben beállítani).

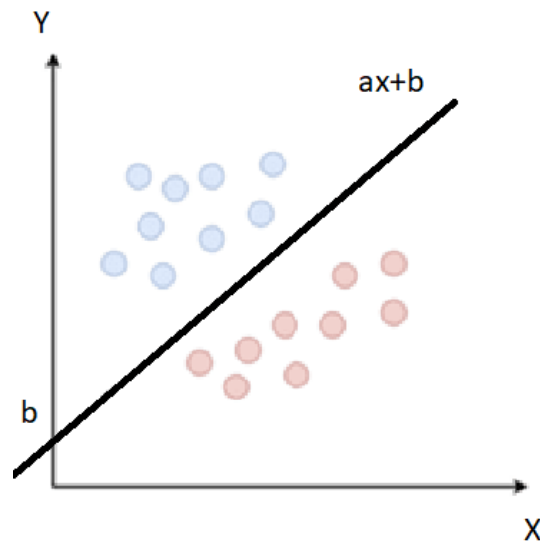


5. ábra: Perceptron modell

A  $b$  Bias érték az eltolást jelképezi, az  $f$  függvény az aktivációs függvény,  $Y$  pedig a kimenet. A perceptron a bemenetére érkező  $X$  értéket megszorozza a bemenethez tartozó  $W$  paraméterrel, több bemenet esetén ezeket összegzi és hozzáadja az eltolás értékét is. Az aktivációs függvény értéke 1, ha a perceptron összeadó kimenete nagyobb, mint 0 és -1 az értéke egyébként. Ezt nem vizsgáljuk most tovább, de az érezhető, hogy bizonyos bemenetekre egy adott neuron érzékenyebb, ha azokhoz nagyobb súly (nagyobb szám) tartozik, vagyis az onnan érkező információkat kiemeli, míg másokat elnyom. [6]

A legegyszerűbb osztályozási feladat látható a 6. ábrán, ahol egy egyenes segítségével a síkot két részre tudjuk osztani. A perceptron modell az 6. ábra szerinti osztályozási feladatot véges lépésben képes megoldani. A kékekkel és pirossal jelölt pontokat egy egyenessel szét lehet választani, azaz a pontok lineárisan szeparálhatóak. Az egyenesnek az általános alakja az  $y=ax+b$  alakban írható le, ahol az „ $a$ ” a meredekség és „ $b$ ” az eltolás paraméter. Egy egyenes leírásához tehát elegendő 2 paraméter megadása és megtanulása. A tanulási folyamat itt azt jelenti, hogy a kék és piros pontokat legjobban szétválasztó egyenest kell megtalálni.





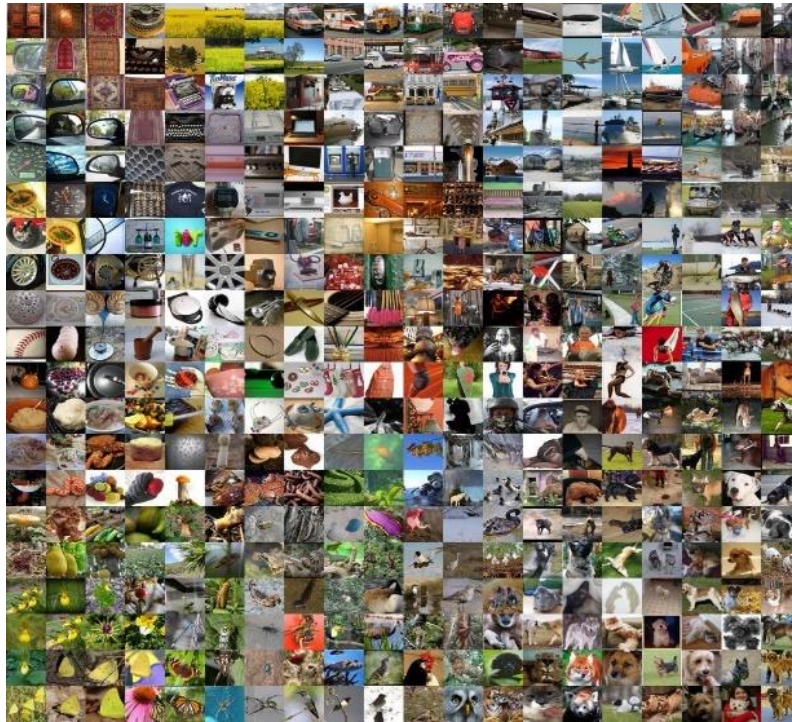
6. ábra: Lineáris szeparáció (szerzői szerkesztés)

A ResNet-50 modellnek a fenti példával párhuzamot húzva nem 2, hanem 23 millió paramétert kell beállítani és megtanulni, amiből sejthető, hogy a fenténél sokkal komplexebb osztályozási feladatok megoldására is alkalmas.

## 3. A MODELL LÉTREHOZÁSA

### 3.1. Tudás importálása

A modellek tanítása költséges és számításigényes feladat, hiszen egy tanulási folyamat alatt a modell paramétereit több ezerszer kell frissíteni és ellenőrizni annak pontosságát. A ResNet-50 modellt az ImageNet kép adatbázisból tanították be. Ez több mint 1,2 millió képből és 1000 osztályból áll. A 8 db nagyteljesítményű Tesla P100-as gyorsítókártya (ezeknek a kártyáknak a piaci ára több millió forint volt a megjelenésekor) felhasználásával a betanítás 29 órát vett igénybe. [5] Ez egy átlagos otthoni felhasználásra készített GPU-n több hétig is eltartott volna, míg hagyományos CPU segítségével teljesen értelmetlen feladat lenne. Kutatók észrevették, hogy előre kiképzett modelleket nagyon jó hatásokkal lehet „átképezni” más hasonló típusú osztályozási feladatra. [7] Ezt a szaknyelvben „transfer learning”-ként említik, én a dolgozatban átképezésnek fogom nevezni az eljárást.



7. ábra: Néhány kép az ImageNet tanítóhalmazból

Az átképzések akkor működnek jól, ha az eredeti tanítóhalmaz és az új tanítóhalmaz elemei valamilyen hasonlóságot mutatnak. A dolgozatban bemutatásra kerülő modell képzésére használt képek utcákat, embereket, járműveket, tüzeseteket tartalmaznak, amik mutatnak hasonlóságot az ImageNet tanítóhalmazában lévő képekhez (7. ábra), ezért az eredeti nagy költségen betanított modellt fel tudtam használni az átképzés során.

### 3.2. A modell képzése

A modell képzéséhez az internetről több ezer képet töltöttem le három kategóriában. A letöltéshez egy speciális scriptet használtam, ami automatikusan a megadott számú képet képes letölteni a keresési kulcsszónak megfelelően. Az egyik kategóriám az épület tűz, amihez 1500, a műszaki mentéshez 1000, az utcaképhez pedig 500 képet tudtam összegyűjteni, azaz összesen 3000 kép képezte az adathalmazomat.



8. ábra: Néhány kép az általam gyűjtött tanítóhalmazból

A teljes adathalmazt szétosztottam két részre, a tanító-halmaz a képek 80%-át, a teszt halmaz a képek 20%-át tartalmazta. A teszthalmazon értékeltem ki a modell teljesítményét, ezeket a képeket a modell nem látta a tanítási folyamat során. A képek előfeldolgozása során az ismétlődő képeket, illetve a kategóriába nem illeszkedő képeket el kellett távolítani.

A modell képzése átképzéssel valósult meg, a ResNet-50 architektúra ImageNet adatkészleten tanított modelljét tanítottam tovább az előbbieken bemutatott képek segítségével. Az átképzési folyamat során a fast.ai programcsomag Python implementációját használtam [8]. Az átképzés néhány percet vett csupán igénybe egy Nvidia Gtx 1060 6Gb-os videokártyán. A ResNet-50 architektúrán túl a ResNet-18 architektúra segítségével is készítettem egy modellt, amely kisméretű egy kártyára integrált néhány dolláros számítógépen (Raspberry Pi) is képes futni.

### 3.3. Az eredmények értékelése

A konvolúciós neurális hálózatok nagyon jól teljesítenek a képosztályozási feladatok során. Megfelelő tanítóhalmaz esetén elegendően robusztusok a különböző bemeneti képekre nézve. A ResNet-50 modell 6%-os hibaarányal teljesített a teszthalmazon, amint az a 9. ábrán látható.



Confusion matrix

	accident	building fire	streets
Actual	accident	building fire	streets
	198	12	8
	2	279	13
	1	5	89
	accident	building fire	streets
	Predicted		

9. ábra: ResNet-50 confusion mátrixa (saját szerkesztés)

A modell az összes tesztkép 6%-át osztályozta rosszul, pl. az épület tűz (building-fire) osztály esetén 279 képet helyesen, 2+13 képet rosszul osztályozott. 2 képet balesetnek jelzett, 13 képet pedig utcaképnek. A rosszul osztályozott képeket érdemes lehet átnézni és elemezni, hogy a hiba miből eredhet.

### 3.4. A modell korlátai

A modell valószínűségi alapon működik, három osztály esetén mindhárom esemény bekövetkezésére (osztályba tartozására) egy 0 és 1 közötti valószínűségi értéket ad meg úgy, hogy a valószínűségi értékek összege 1. (pl. a 0,3 valószínűségi érték 30%-os valószínűséget jelent) A modell döntése tehát értelmezés kérdése. Sok osztály esetén (több száz kategória) előfordulhat, hogy két vagy több osztálynak nagyon hasonló a bekövetkezési valószínűsége, és a valószínűségi érték is alacsony (pl. 3-5 %) azaz a modell bizonytalan a döntésben. Ez azt jelenti, hogy a modell nem ismeri fel az adott képet, nem tudja azt osztályozni.

A modell sohasem fogja elérni a 0%-os hibarányt a valóságban. Lehet a tesztképeknek olyan halmaza, ahol esetleg előfordulhat ilyen, de a világunk folyamatosan változik és előfordulhatnak olyan esetek, amelyek a tanítóhalmazban egyáltalán nem vagy csak kis

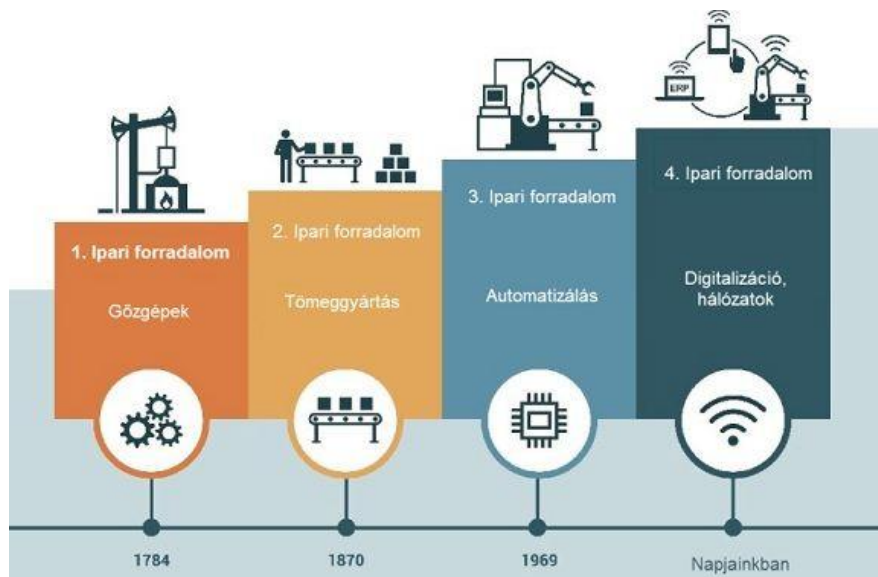


számban fordultak elő. Ilyen esetknél a modell bizonytalan lesz a döntését illetően. A Tesla autók önvezető rendszere is használ konvolúciós hálózatokat, és bizony vannak olyan esetek, amikor a rendszer nem ismer fel valamilyen ritkán előforduló eseményt. Az önvezető autóknál ilyenkor előfordulhat, hogy egy baleset lesz a kimenetele egy rossz döntésnek.

A modellt fel kell készíteni különböző fényviszonyokra és időjárási helyzetekre, hiszen amennyiben a modell gyengébb fényviszonyok (szürkület, éjszaka) között készült fényképet kell értékeljen, de a tanítóhalmazban csak tökéletes napos időben készült képek találhatóak, a modell nem fogja hozni az elvárt eredményeket.

### 3.5. Az adat az új olaj

Egy mondás szerint „az adat az új olaj” vagy „az adat az új arany”. A 10. ábrán az ipari forradalmak láthatóak. Az első ipari forradalom hajtó ereje a szén volt, a másodiké az olaj, a harmadik ipari forradalom során az elektronika, a nyomtatott áramkörök bírtak a legnagyobb jelentőséggel. Jelenleg a negyedik ipari forradalom korát éljük, amiben a mesterséges intelligencia megjelenésével a legnagyobb hajtóerő az adat, hiszen a mesterséges intelligencia rendszerek képzéséhez nagyon sok jó minőségű adatra van szükség.



10. ábra: Ipari forradalmak

A dolgozatban nem vállalkoztam arra, hogy olyan modellt építsek, ami minden körülmények között megbízható működést garantál, hiszen ehhez nagyon sok adatra (esetünkben



felcímkézett fényképre) lenne szükség, ami nem állt rendelkezésemre. Az internetről letöltött adatok viszont nagyon sokfélék, változatosak, ezért ezen adatokból épített modell lehet az alapvonal, ami meghatároz egy indulási szintet, amit aztán további adatok beszerzésével bővíteni, fejleszteni lehet. A modell építésénél az igazi értéket már nem a számítási kapacitás (hardver erőforrás), nem is az architektúra (szoftveres módszertan) képviseli, hiszen ezek még a 3. Ipari forradalom termékei, hanem maga a feldolgozandó adat.

## 4. HOGYAN MŰKÖDHET EZ A GYAKORLATBAN?

A következőkben megvizsgáljuk az előbbieken felvázolt elméleti dolgok gyakorlati kivitelezésének lehetőségét az 1. ábra alapján.

Ahhoz, hogy a gyakorlatban használni tudjuk a modellt a 4. fejezetben bemutatott módszer segítségével, először be kell azt tanítani. A tanítást csak egyszer kell elvégezni, esetleg később több adat birtokában a modellt lehet finom hangolni a még pontosabb működés eléréséhez.

### 4.1. IP kamerák

Az információ a „földön hever”, körül vesz minket, el kell kapni azt. Ezt esetünkben kamerák segítségével tudjuk megvalósítani. A kamerák éjjel-nappal folyamatosan nagyfelbontásban közvetítik az adatfolyamot. A modern kamerák általában IP (Internet Protocol) rendszerűek, ami annyit tesz, hogy önállóan egy hálózatba szervezve képesek működni [9]. A hálózat lehet intranet (belső zárt), vagy akár internet alapú. Mindkét típusú hálózat képes a képfolyamot továbbítani a hálózat egy másik pontjára, ahol azt fel tudjuk dolgozni. Ezt az Internet segítségével tudjuk megtenni. Sok esetben a kamera képe nyilvános, ilyenkor csak azt a webcímet kell ismerjünk, ahova a kamera közvetíti a képét. A kamera másodpercenként akár 30 képkockát is közvetíthet, ami jelentős adatmennyiséget képvisel.



## 4.2. A tűztorony feldolgozó egysége

A kamerák képeit egy feldolgozó egységbe vezetjük, ami egy nagyteljesítményű szerver számítógépet jelent a gyakorlatban. Ennek a szervernek akár több száz kamera által biztosított képfolyamot (videó jel) kell feldolgozni, amihez óriási számítási kapacitás szükséges. Ez a hatalmas számítási kapacitás csak abban az esetben szükséges, ha az időbeliség kulcsfontosságú. Esetünkben, ha csak pl. 10 másodpercenként egy képkockát használunk fel, az is elegendő adatsűrűséget jelent. Ez a gyakorlatban azt jelenti, hogy 10 másodpercenként ellenőrizzük, hogy egy adott kamera által lefedett területen történt-e tüzeset. Ezzel a trükkel elérhető, hogy egy átlagos teljesítményű szerver számítógép is jelentős számú kamera képét legyen képes feldolgozni. A gyakorlatban egy NVIDIA GTX 1080 GPU egységgel felszerelt számítógép kb. 100 képet tud feldolgozni másodpercenként. Ez egy átlagos videófolyam (25 képkocka másodpercenként) esetén 4 kamera képének feldolgozását jelenti. Amennyiben minden kamerából 10 másodpercenként csak egy képkockát dolgozunk fel, úgy  $100 \cdot 10$ , azaz 1000 kamerát jelentene elméletileg. A gyakorlatban ez a szám biztosan kisebb, de jelentősen több a 4 kameránál.

A képek feldolgozása azt jelenti, hogy a képet „megmutatjuk” a modellünknek, majd a modell az adott képre elvégzi az osztályozást és egy valószínűségi eredményt ad a kimenetén mind a három (tűz, közlekedési baleset, normál utcakép) osztályra vonatkozóan. A képek feldolgozása, „áttolása” a neurális hálózaton azért időigényes feladat, mert amint azt a korábbiakban láthattuk, egy-egy architektúra több millió paraméterrel rendelkezik, amelyekkel számításokat kell végezni, ez a számítási idő esetünkben kb. 10 ms képenként.

## 4.3. A modell kimeneti értékeinek értelmezése

A valószínűségi kimeneteket értékelni kell, hogy továbbítjuk-e a 112-es rendszer, vagy egyéb tűzátjelző központ felé a jelzést. Miután a képet megmutattuk a modellnek, annak kimenetén az osztályok számának megfelelő, esetünkben három számot kapunk. Ezek a számok az egyes osztályokhoz való tartozás valószínűségét jelölik. A valószínűségek 1-nél kisebb számok és a valószínűségek összege egy. A 11. ábrán láthatunk egy bemeneti képhez tartozó kimeneti értékeket.



('building fire', tensor(1), tensor([6.2518e-08, 9.9989e-01, 1.1086e-04]))

## 11. ábra: Valószínűségek értelmezése

A modell egy három elemű tensor (vektor) kimenetet adott meg normál alakban. A 0. osztályhoz (műszaki mentés)  $6,25 \cdot 10^{-8}$ , ami 0.0000000625, az 1. osztályhoz (épülettűz)  $9,9989 \cdot 10^{-1}$ , ami 0,99989, a 2. osztályhoz (utcakép) pedig  $1,1086 \cdot 10^{-4}$ , ami 0,00011086 valószínűségek tartoznak. Ezeket százalékosan kifejezve: műszaki mentés 0,00000625%, épülettűz 99,989% és utcakép 0,011086%.

A modell tehát meglehetősen magabiztosan, épülettűznek értékelte a bemenetként kapott képet. A kép egy magyarországi tüzeset során készült.





('streets', tensor(2), tensor([1.1338e-05, 4.2500e-05, 9.9995e-01]))

## 12. ábra: Valószínűségek értelmezése

A 12. ábrán egy budapesti utcakép látható, amelyet a modell nagy magabiztossággal 99,9% utcaképnek értékelt.

### 4.4. Mit figyel a modell egy képen?

A modell belső működése elvont és nehezen érthető, hiszen a képek feldolgozása során a képek átalakulnak egy másfajta reprezentációba. Ezt úgy kell elképzelni, hogy az egyes feldolgozási lépések során a kép felbontása csökken (fizikai kiterjedése magasság és szélesség), azonban egyre több réteg jelenik meg, amelyek különböző alakzatokat és egyre bonyolultabb formákat reprezentálnak. Esetünkben pl. megjelennek olyan rétegek, amelyek a tűzre aktiválódnak, tehát ha a képen megjelenik valahol egy tűz ez a réteg aktiválódik.

A modell jobb megértéséhez használják a Gradient-weighted Class Activation Mapping eljárást, amelynek az a lényege, hogy a képen vizuálisan is meg tudjuk jeleníteni azokat a részeket, amely alapján a modell döntött az osztályozás során [10]. A 13. ábrán a világosabb sárga színek jelzik azokat a részeket, amely miatt a modell az aktuális osztályt választotta. A baloldali képen a modell az erős füstképződés miatt döntött a tűz kategória mellett.



### 13. *ábra: Aktivációs térképek (saját szerkesztés)*

Viszont a jobb oldali képen a modell utcaképnek értékelte a képet. A döntésében a bokrok a házfal ablakokkal és az égbolt volt hatással a kimenetre. Ezt a fajta ábrázolást a rosszul kategorizált képeknél is fel lehet használni annak a vizsgálatára, hogy a modell miért hozott egy adott döntést.

Egy városi legenda szerint katonai célokra olyan modellt fejlesztettek, ami a képeken tankokat ismert fel. Laboratóriumi körülmények között a modell jól teljesített, azonban a valóságban teljesen használhatatlan volt. Kiderült, hogy a tankokat ábrázoló fotók speciális körülmények között készültek, pl. mindegyik ragyogó napos időben készült, így a modell nem a tankok jelenlétét tanulta meg, hanem azt, hogy milyen idő van.

A történet arra világít rá, hogy a modell működését nagyon aprólékosan kell ellenőrizni, illetve a tanulóhalmaz kiválasztása során az adatokat gondosan kell megválasztani.

#### **4.5. Az eredmények továbbítása**

Miután a modell meghozta a döntését az eredményt továbbítani kell a tűzátjelző központ felé. Itt nagyon fontos, hogy nem csupán a riasztási jel továbbítását kell megoldani, hanem azt a képet is, amely generálta a riasztást. A kép alapján egy ember már könnyedén el tudja dönteni, hogy az esethez valóban riasztani kell tűzoltóegységeket vagy esetleg téves a jelzés. A kamera



azonosítója alapján élőképet kérhet a helyszínről, ami további segítséget nyújthat az esemény minősítéséhez.

Fontos kiemelni, hogy egy ilyen intelligens rendszer kiválthat sok humán erőforrást, akik a kamerákat figyelnek, azonban nem szabad túlértékelni a modell tudását, azt mindenképpen humán kontroll alatt kell tartani, mintegy segítve a műveletirányítási feladatot ellátókat.

## 5. TOVÁBBI FELHASZNÁLÁSI LEHETŐSÉGEK

### 5.1. Életmentés

A kameraképek alapján nemcsak baleseteket, tüzeket lehet megfigyelni, hanem pl. egy huzamosabb ideje földön fekvő embert is tud azonosítani a rendszer, ilyenkor másfajta működési mechanizmusok lépnek életbe és az adatlap a 112-es rendszeren már a mentőszolgálat irányába közlekedik.

### 5.2. Rendvédelem, biztonság

A keresési algoritmusoknak csak a fantázia szabhat határt, illetve a rendelkezésre álló tanítóadatok. A rendszer képes lehet pl. fegyverek, kések, maszkos emberek felismerésére, majd riasztási jel vagy élőkép kapcsolására a megfelelő értékelő központokba.

### 5.3. Speciális terek tűzvédelme

Gyakran vannak olyan helyszínek, ahol nagymennyiségű - akár éghető - alapanyagot tárolnak, de csak hagyományos érzékelők állnak rendelkezésre. Amennyiben egy nyitott térrészben helyezkednek el az éghető anyagok, úgy a hagyományos érzékelők csak korlátozottan alkalmazhatóak, hiszen előfordulhat, hogy sem a hőfejlődés, sem a füstképződés nem képes az érzékelőket inicializálni. Ezekbe a terekbe fejlesztették ki a hőkamerás megfigyelési eszközöket, azonban bizonyos korlátozások élnek, hiszen egy nem elektromos targonca, vagy egyéb robbanómotoros jármű téves jelzést eredményezhet. Ezekben az esetekben a kamerás megfigyelés akár csak kiegészítő elemként is javíthatja a tűzjelzés pontosságát.



## 5.4. Erdők tűzvédelme

A közelmúlt nagy nemzetközi erdőtüzei rávilágítottak, hogy a természeti értékekben a tűz olyan méretű pusztítást tud végbevinni, aminek következményei nem vagy, csak részben visszafordíthatóak. Az erdő és vegetációtüzek időbeni felderítése kulcsfontosságú a sikeres beavatkozás érdekében. Az idő előrehaladtával négyzetes arányban nő a leégett terület nagysága. Természetes vagy épített magaslati pontokon kamerák elhelyezésével könnyebben felderíthetőek lennének a tüzesetek. Az éjszakai észleléshez persze kifinomultabb kamerák kellene, de korlátozott képességekkel működhetne a megfigyelés.

## 5.5. Veszélyes áru szállítványok

A veszélyes áruk közúti szállítását kezelhetjük kockázatként egy útszakaszra vonatkozóan. Egy adott közút veszélyességét sok szempont határozza meg például az azon szállított veszélyes áruk mennyisége, időbeli eloszlása, a forgalom nagysága a különböző napszakokban. Kamerák segítségével a közúti szállításban közlekedő jelölt ADR-es szállítványok azonosíthatóak lennének, sőt az UN szám, illetve a jármű nagysága alapján becsülhető lenne az adott útszakasz veszélyes áru forgalma. Az ebből épített adatbázis segíthetné az ellenőrzések tervezését, hiszen a hét napjai és a napszakok szerinti felbontásban lehetne összegezni az információkat.

## 5.6. Kontextusok megfigyelése

A dolgozatban bemutatott konvolúciós neurális hálózati modell csak egyszerű osztályozási feladatokhoz lett kifejlesztve, ahol egy kép egy kategóriát jelöl. A valóság azonban általában ennél bonyolultabb, egy képen több különböző objektum is azonosítható, pl. emberek, autók, épületek, stb. Az R-CNN architektúra család, illetve ennek tovább fejlesztett változatai pl. a legújabb Detectron2, képesek az objektumok azonosításra egy képen belül és azokhoz egy körülíró téglalapot rendelnek, de akár a képek pixel szintű osztályozása is lehetséges MaskR-CNN architektúrák használatával [11].

Ezek a fejlett architektúrák képesek kontextusba helyezni a megfigyeléseket, egy apró íves tárgyról szürke háttérrel nem tudjuk eldönteni, hogy mit ábrázolhat, azonban, ha látszik a tárgy



környezete egy tengerpart, ahol emberek vannak (14. ábra) a kontextusból felismerjük, hogy a képen apró sárkányrepülőket eregetnek. Ezek a kontextusok a modellt rendkívül erőssé tehetik és jelenleg is aktív kutatások folynak ezeknek a lehetőségeknek a kihasználására.



14. ábra: R-CNN objektum detektálás

A képeken a lángok, a füst és az autó vagy épület felismerésével a riasztást lehetséges finomítani. Egy ilyen modell képes lehet arra, hogy leíró jellegű adatokat is adjon a riasztás mellé, pl. a képen 10 ember, 3 sérült gépkocsi és lángok láthatóak. Ezen információk nagyban segíthetik a műveletirányítás munkáját.

## 6. ÖSSZEFOGLALÁS

A tüzesetek felderítésének régi új módszerét mutattam be. Az elvek a régié, azonban a módszerek már a 21. század legújabb eredményeire építenek. A tüzesetek egyik felderítési módszere a vizuális észlelés, amely az évszázadok során nem sokat változott, de a mesterséges intelligencia megjelenésével már a „számítógépek látása” is elérhető és új dimenziókat nyithat. A dolgozatban bemutatam a gépi látás felhasználását osztályzási feladatok megoldására. Áttekintettem a gépi tanulási modell elméleti háttérét, hiszen nagyrészt egyszerű számításokra támaszkodik, de rendkívül bonyolult feladatok megoldására is képes. Minden



modell egy matematikai függvényillesztési feladatnak tekinthető, ami során a függvény paramétereit „kell csak” beállítani. A modell tanítása során megismerkedtünk az adatok kiemelt szerepével és értékével, majd a modell korlátaival is. Ezt követően bemutattam a tűztorony gyakorlati megvalósítását, amely kamerák sokaságának képeit dolgozza fel és elemzi ki tűzvédelmi szempontból. A megvalósított modell képes felismerni az épületek tüzeseteit és a közlekedési baleseteket 94%-os valószínűséggel, az általam összegyűjtött tesztképek halmazán. Ezt az információt pedig továbbítva egy tűzátjelző központba megvalósulhat a tüzesetek automatizált felismerése és átjelzése kameraképek alapján, amely a fő célkitűzésem volt a dolgozat készítésekor. A dolgozat végén bemutattam a további, főként katasztrófavédelmi célú felhasználási lehetőségeket és a modell további fejlesztési lehetőségeit, amely segítségével kifinomultabb jelzések generálhatóak. Ezek a jelzések már tényleg egy intelligensnek tűnő bejelentőt engednek felsejleni, pedig a háttérben nincsen más csak számok és számok.

## FELHASZNÁLT IRODALOM

[1] [https://hu.wikipedia.org/wiki/Big\\_data](https://hu.wikipedia.org/wiki/Big_data)

[2] <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>

[3] <https://www.kdnuggets.com/2016/11/intuitive-explanation-convolutional-neural-networks.html/3>

[4] Shankar, V., Roelofs, R., Mania, H., Fang, A., Recht, B. & Schmidt, L.. (2020). Evaluating Machine Accuracy on ImageNet. Proceedings of the 37th International Conference on Machine Learning , in Proceedings of Machine Learning Research 119:8634-8644

[5] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in CVPR, 2016.



- [6] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>
- [7] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: *Advances in neural information processing systems*, vol 27; 2014.
- [8] <https://github.com/fastai/fastai>
- [9] <https://hu.wikipedia.org/wiki/Internetprotokoll>
- [10] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization.. *ICCV* (p./pp. 618-626), : IEEE Computer Society. ISBN: 978-1-5386-1032-9
- [11] Ren, S., He, K., Girshick, R. B. & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama & R. Garnett (eds.), *NIPS* (p./pp. 91-99),

## ÁBRAJEGYZÉK

1. ábra Tűztorony modell

Forrás: készítette a szerző

2. ábra Konvolúciós neurális hálózat

Forrás: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>

3. ábra Konvolúciós neurális hálózat rétegeinek vizualizációja

Forrás: <https://www.kdnuggets.com/2016/11/intuitive-explanation-convolutional-neural-networks.html/3>

4. ábra ResNet34

Forrás: K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016



5. ábra Perceptron modell

Forrás: <https://jameskle.com/writes/neural-networks-101>

6. ábra Lineáris szeparáció

Forrás: készítette a szerző

7. ábra Néhány kép az ImageNet tanítóhalmazból

Forrás: <https://cs.stanford.edu/people/karpathy/cnnembed/>

8. ábra Néhány kép az általam gyűjtött tanítóhalmazból

Forrás: DuckDuckGo.com

9. ábra ResNet-50 confusion mátrixa

Forrás: készítette a szerző

10. ábra Ipari forradalmak

Forrás: <https://digitalizationindustry.com/hu/2017/07/26/a-negyedik-ipari-forradalom/>

11. ábra Valószínűségek értelmezése

Forrás: Pécsi Hivatásos Tűzoltóság képtárháza

12. ábra Valószínűségek értelmezése

Forrás: Goole utcakép

13. ábra Aktivációs térképek

Forrás: készítette a szerző

14. ábra R-CNN objektum detektálás

Forrás: <https://towardsdatascience.com/faster-rcnn-object-detection-f865e5ed7fc4>

**Karsa Róbert** tűzoltó ezredes, igazgatóhelyettes

Baranya Vármegyei Katasztrófavédelmi Igazgatóság

robert.karsa@katved.gov.hu

ORCID ID: 0000-0003-0502-1508