[Külső fekete borítólap formátuma]

Széchenyi István Egyetem

Gépészmérnöki, Informatikai és Villamosmérnöki Kar

Informatika Tanszék

**SZAKDOLGOZAT**

**Németh Valentin**

**Mérnök Informatikus BSc**

2021

|  |
| --- |
| Németh Valentin, 2021 |

**SZAKDOLGOZ****AT**

**Kotta digitalizáló szoftver fejlesztése**

**Németh Valentin**

**Mérnök Informatikus BSc**

**2021**

# Nyilatkozat

Alulírott, Németh Valentin (KNPDXS) Mérnök Informatikus BSc szakos hallgató kijelentem, hogy a Kotta digitalizáló szoftver fejlesztése című szakdolgozat feladat kidolgozása a saját munkám, abban csak a megjelölt forrásokat, és a megjelölt mértékben használtam fel, az idézés szabályainak megfelelően, a hivatkozások pontos megjelölésével.

Eredményeim saját munkán, számításokon, kutatáson, valós méréseken alapulnak, és a legjobb tudásom szerint hitelesek.

Győr, [dátum]

hallgató

# Kivonat

Kotta digitalizáló szoftver fejlesztése

A szakdolgozat célja, hogy elkészítsek egy zenei kotta digitalizálását megvalósító szoftvert. A szoftver bemenetként kap egy képet egy zenei kottáról, amelyen a hangjegyek és egyéb szimbólumok találhatóak. A program felismeri ezeket, valamint az elhelyezkedésüket is, ez alapján feldolgozza a képet, melyet egy olyan kimenetté alakít át, amelyet egy kottaolvasásra alkalmas program is értelmezni tud bemenetként.

A dolgozat első felében a szoftver fejlesztésének elméleti hátterével foglalkozok részletesebben. Ebben a részben ismertetem a legfontosabb alapokat, melyek a szoftver tényleges fejlesztéséhez elengedhetetlenek. Előszőr inkább általánosan, a használt programnyelvről és a megvalósításhoz szükséges információkról lesz szó, majd utána specifikusabban a szakdolgozathoz szükséges elemekről.

A diplomamunka második felében bemutatom, hogyan valósítottam meg a szoftver elkészítését. Itt már főként a gyakorlati megoldások lesznek bemutatva, a használt programkódokkal alátámasztva. Lépésről lépésre követhető lesz a programozás menete, s mellette folyamatos magyarázatok segítik az olvasót annak megértésében, hogy a feladatot miért az adott metódusokkal oldottam meg. A megvalósítás végén keletkező kimenetet is bemutatom, valamint különböző tesztelésekre is sor fog kerülni.

A szakdolgozat elkészítéséhez különböző programozási ismeretek, valamint a gépi látás és az azt támogató programcsomagok szükségesek. A megvalósítás során sok új ismeretre tettem szert, amelyek az élet során fontos tudást jelenthetnek számomra, amennyiben a jövőben szoftverfejlesztési feladatokkal lesz dolgom.

# Abstract

Developing a sheet music digitizer software

The aim of the dissertation is to create a software for digitizink a sheet music. The software gets a sheet music as an input, that includes notes and other symbols. The program recognizes these as well as their location, it processes the image, which it converts into an output that can also be interpretes an an input by a program that is capable of reading sheet music.

In the first half of the dissertation i deal with the theoretical background of software development in more detail. In this section, I describe the most important basics, that are essential for the actual development of the software. First, I will be more general, with the presentation of the programmin language used and the information needed for the implementation, then the more specific elements, whicg are needed for the dissertation.

In the second half of the dissertation I will show you how I implemented the software. Here, mainly the practical solutions will be presented, supported by the program codes used. The programming process will be followed step by step, and continuous explanations will help the reader to understand why I solved the task with the given methods. I will also present the output at the end of the implementation, and various test will take place.

The preparation of the dissertation requires various programming skils, as well as machine vision and the software packages that support it. During the implementation, I gained a lot of new knowledge, which can be an important knowledge for the rest of my life, if I have to deal with software development tasks in the future.

# Tartalomjegyzék

Tartalom

[Nyilatkozat 3](#_Toc86750418)

[Kivonat 4](#_Toc86750419)

[Abstract 5](#_Toc86750420)

[Tartalomjegyzék 6](#_Toc86750421)

[1. Bevezetés 7](#_Toc86750422)

[2. A programozási nyelv 8](#_Toc86750423)

[2.1 A nyelvek fejlődése 8](#_Toc86750424)

[2.2 A programozási nyelvek generációi 8](#_Toc86750425)

[2.3 Imperatív és Deklaratív nyelvek 10](#_Toc86750426)

[2.4 Eljárásorientált és Objektumorientált nyelvek 10](#_Toc86750427)

[3. A Python 11](#_Toc86750428)

[3.1 Története 11](#_Toc86750429)

[3.2 A Python néhány fontos jellemzője: 12](#_Toc86750430)

[3.3 További tulajdonságok 13](#_Toc86750431)

[3.4 Milyen programok írásához jó? 13](#_Toc86750432)

[4. Gépi látás 14](#_Toc86750433)

[4.1 Alkalmazási területei 15](#_Toc86750434)

[5. A program megtervezése 17](#_Toc86750435)

[5.1 Thresholding 18](#_Toc86750436)

[5.2 Vonalkeresés 20](#_Toc86750437)

[5.2.1 Hough Line Transform 20](#_Toc86750438)

[5.2.2 Probabilistic Hough Transform 22](#_Toc86750439)

[5.3 Kontúrkeresés 24](#_Toc86750440)

[5.4 Konvolúciós neurális hálózat 26](#_Toc86750441)

[5.4.1 A hagyományos CNN szerkezete 28](#_Toc86750442)

[5.4.2 Objektum felismerés a CNN használatával 29](#_Toc86750443)

[5.5 Régiókon alapuló konvolúciós neurális hálózat 30](#_Toc86750444)

[6. A program megvalósítása 32](#_Toc86750445)

[Irodalomjegyzék 33](#_Toc86750446)

# Bevezetés

Szakdolgozatom célja, hogy elkészítsek egy zenei kotta digitalizálására alkalmas szoftvert. A megvalósítás során a fejlesztés lépéseit és az elméleti hátterét is dokumentálom. A szoftver Python programozási nyelvben lesz fejlesztve, különböző gépi látást támogató csomagok segítségével. A program bemenetként egy képet kap a kottáról, amely lehet digitális kotta, vagy fotó egy papír alapú kottáról. A szoftvernek fel kell ismernie a zenei kotta meghatározott alkotóelemeit és azoknak elhelyezkedését, majd a felismert kész képből valamilyen digitalizált formában egy olyan kimenetet generál, amelyet egy kottaolvasó szoftver is képes felismerni. A megvalósítás során kiválasztom a megfelelő metódusokat a felismerni kívánt elemek definiálására, generálok egy olyan kimenetet, amelyet egy kottaolvasó szoftver is kezelni tud, valamint különböző teszteket készítek, amelyeken lefuttatom a programomat, hogy az esetleges hibákat ki tudjam küszöbölni. A tesztek különböző bemeneti képek lesznek, amelyeken a szoftver hatékonyságát is ellenőrzöm.

# A programozási nyelv

Minden programozási feladat első lépése meghatározni, hogy milyen környezetben akarjuk megvalósítani a fejlesztést. Ezt legnagyobb mértékben a választott programozási nyelv határozza meg. Manapság már sokféle lehetőségből választhat a fejlesztő, azonban ez nem mindig volt így.

## A nyelvek fejlődése

A programozás, mint fogalom az 1950-es években került először szóba, miután 1952-ben Neumann János elvei alapján elkészült az EDVAC, az első programozható számítógép. Itt még nem a mai értelemben vett programozást kell érteni, hiszen az akkori programok még csak matematikai problémák megoldására voltak alkalmasak. A kódok megírása és maga a kódnyelv is teljesen kezdetleges volt még abban az időben, a programozás tulajdonképpen nem volt más, mint a számítógépnek megadott utasítások sorozata. A programok „gép-függőek” voltak, azaz egy programot egy adott számítógépre lehetett csak megírni. A kódok megírása hosszú és körülményes munka volt. [2]

A 60-as, 70-es években sikerült gyorsabbá és egyszerűbbé tenni a programfejlesztési módszereket, és megszülettek az első magas szintű programozási nyelvek is. Az új programozási nyelveknek köszönhetően olyan módszerek és technikák alakultak ki, amelyek nagyban megkönnyítették a szoftverfejlesztők feladatát. Az újonnan kialakult nyelvek között voltak általános célúak, de voltak specializáltak is. [2]

A 80-as, 90-es években jelent meg az objektumorientált programozás, valamint a Script-nyelvek, amelyek ma is a leghasználatosabb módszerek. Egyes scriptnyelveket hibridnyelvnek is neveznek, mivel egyszerre van lehetőség bennük az eljárásorientált és az objektumorientált programozásra is. Ilyen például a PHP, a Javascript vagy az Actionscript. [2]

## A programozási nyelvek generációi

1. generáció:

* gépi kódban írták őket
* nehézkes és körülményes programírás
* erősen gépfüggő
* gyors lefutási idő
* a processzor közvetlenül végrehajtja az utasításokat
* az Assembly nyelv is ide tartozik, de az már szöveges kódú
* Assembler, azaz fordítóprogram szükséges hozzá
* kapcsolatban áll a gépi nyelvvel

1. generáció:

* megjelentek a magasszintű programnyelvek
* áttekinthetőbb a program
* egyszerűbb és gyorsabb programírás
* hosszabb lefutási idő a fordító miatt
* kevesebb hibalehetőség a fordítók fejlettsége miatt
* kialakulnak az alapvető utasítástípusok
* megszületnek az elemi adattípusok
* létrejönnek az első összetett adatszerkezetek, a tömb és a rekord
* pl. FORTRAN, ALGOL, COBOL, BASIC

1. generáció

* a nyelvek elkezdtek a problémák alapján specializálódni
* megjelent a procedurális szemlélet és a struktúrák hangsúlyozása
* megszülettek az általános célú, az objektumorientált és a specializált nyelvek
* az általános célúak közé tartozik a PASCAL, ADA, APL, C
* az objektumorientált nyelvekhez tartozik a C++, Turbo Pascal, Smalltalk
* a specializált nyelvek többek között a Logo, Forth

1. generáció:

* 4GL rendszerek
* ezek már kifejezett programnyelvek
* magasszintű programnyelvre épülő, komplex, objektumorientált fejlesztői környezet
* modularitás jellemzi
* adatbázissal szorosan kapcsolódnak
* pl. Visual Basic, Visual C++, Java, Object Pascal

[2]

## Imperatív és Deklaratív nyelvek

A programnyelveket a kód szerkezete alapján csoportosíthatjuk imperatív és deklaratív csoportba. Ezek néhány lényeges dologban térnek el egymástól.

Az imperatív nyelvek algoritmikus nyelvek, ahol a program főként utasításokból áll, és a változók a legfontosabb részei a programoknak. Itt a programozó határozza meg a feladat megoldását, és a tárkezelésről is ő gondoskodik. Ezeknek a programozási nyelveknek két alfajtája van: az eljárásorientált és az objektumorientált. [2]

A deklaratív nyelvek esetében a programozó csak a megoldandó feladatot deklarálja, de a megoldást nem ő dolgozza ki. Ugyan változók ennél a programnyelv típusnál is megtalálhatóak, jelentésük más, mint az imperatív nyelveknél. Hasonlóan a másik nyelvhez, itt is utasításokat ad a szoftverfejlesztő a számítógépnek, ám a memóriát nem tudja kezelni. A deklaratív nyelvek alfajtái a funkcionális és a logikai nyelvek. [2]

## Eljárásorientált és Objektumorientált nyelvek

Az eljárásorientált nyelvek egyik ismertető jele, hogy a megoldandó feladatot kisebb részegységekre, más néven eljárásokra bontja. Ezeket az alprogramokat jól körül kell határolni, mivel a fő program majd ezeket fogja sorozatosan meghívni. Amikor megtörténik a meghívás, az eljárás kap egy vagy több bizonyos paramétert, amelyet a benne meghatározott logika szerint valamilyen visszatérési értékké alakít át és ad vissza a fő programnak. A kettő legfőbb adattípusa az eljárásorientált programoknak a tömb és a rekord. [2]

Az objektumorientált programnyelvek - ahogy a nevükben is benne van – objektumokon alapulnak. Objektumnak hívjuk a programban szereplő adatokat és a hozzájuk tartozó műveleteket. Mindegyik objektum rendelkezik tulajdonságokkal és metódusokkal. Ezek a nyelvek osztályokra épülnek, és az objektumok az osztályok egy-egy példányai. Az osztályoknak lehetnek ősosztályaik, akiktől örökölhetnek tulajdonságokat, metódusokat, illetve lehetnek származtatott osztályaik, akiknek ők adhatnak tovább. Az öröklődésen kívül még az egységbezárás és a polimorfizmus miatt is jelentősen eltér az eljárásorientált nyelvektől. [2][9]

# A Python

A szoftver elkészítéséhez a Python programozási nyelvre esett a választásom. A Python egy rugalmas, egyszerű kódú, dinamikus, objektumorientált és általános célú, magasszintű programozási nyelv, amely széles körben alkalmazható. Ingyenesen használható, nyílt forráskódú, és nagy népszerűségnek örvend, mind a vállalatok, mind a magánszemélyek körében. [12]

Ez a nyelv, különböző programozási stílusokat is támogat, például az objektumorientált illetve a strukturális. Rugalmasságának egyik fő oka, hogy más nyelven írt programokat is képes használni, amiket modulokként importál a saját környezetébe. A beimportált programokat aztán tovább lehet bővíteni, alakítani. [12]

## Története

A Python fejlesztése az 1990-es évek elején indult Hollandiában, ahol Guido Van Rossum a CWI alkalmazottja 1991-ben bemutatta a Python első verzióját. A projekt célja eredetileg az ABC programnyelv helyettesítésére szolgáló nyelv kifejlesztése volt. 1995-ben megjelent a Python 1.2 még szintén a CWI fejlesztésével. [12]

2000-ben elkészült a Python 1.6, majd ezután Guido Van Rossum összeállt a BeOpem.com-al, és megalapították a Python Laboratoriest. Még ugyanabban az évben a Python Laboratories kifejlesztette a Python 2.0-t. A 2.0-ban már olyan újdonságok voltak, mint a listák használata és a keletkezett memóriaszemét számolása. [12]

A 2.0-s verzió kiadása után megalakult a Python Software Foundation, amely a mai napig tulajdonosa és fejlesztője az immár ingyenes és nyílt forráskódú Pythonnak. 2008-ban kiadták a Python 3.0-t, amely ismét jelentős változásokon esett át. Bevezették az integer típusokat, új szintaxisokat hoztak létre, és megváltoztattak régieket, létrejött a print funkció, valamint ez a verzió szándékosan nem kompatibilis a korábbi verziókkal. [12] [14]

## A Python néhány fontos jellemzője:

* gyors és hatékony:
  + Tartalmaz egy standard belső könyvtárat. Minden olyan lehetőséget biztosít, amely a programozáshoz szükséges, az alapműveletektől a speciális funkciókig. Ezek a harmadik féltől származó eszközök mindent lehetővé tesznek a Pythonban. Például egy webszerver létrehozásához elég három sornyi kód megírása. [12]
* más technológiákat is támogat:
  + Többek között támogatja a COM és a .Net objektumokat. Ezen kívül kiegészítőket és alternatívákat is létrehoztak a Python számára, amik megkönnyítik az integrált módban történő munkát az objektumokkal. [12]
* operációs rendszer független:
  + A Python scriptek különböző operációs rendszereken is használhatóak, például Windows, Linux, MacOS. Néhány verzió kiadásra került a .Net, Java és Nokia S60 használatra is. A programok kimenete az összes platformon igencsak hasonló. [12]
* egyszerű:
  + Ez egy nagyon magas szintű nyelv, amely a sok forrásnak köszönhetően könnyen megtanulható. A harmadik féltől származó eszközök sokfélesége miatt a felhasználókat folytonos tanulásra és fejlesztésre ösztönzi. [12]
* nyílt forráskódú:
  + Annak ellenére, hogy minden jog a Python Software Foundation kezében van, nyílt a forráskódja, és semmilyen korlátozáshoz nincs kötve a használata, terjesztése vagy a módosítása. [12]

## További tulajdonságok

A Python egy több paradigmás programnyelv: az objektumorientált programozás és a strukturált programozás teljes mértékben támogatott, és számos olyan nyelvi funkció létezik, amelyek támogatják a funkcionális programozást és a szempontorientált programozást. Sok más paradigma támogatott a kiterjesztések használatával, ideértve többek között a logikai programozást. [12]

A Python a memóriakezelést a dinamikus gépelés, a referenciaszámlálás és a ciklusérzékelő szemétgyűjtés kombinációjával oldja meg. Egy másik fontos tulajdonság a dinamikus névfeloldás, ami a metódus és a változó nevét a program végrehajtása során köti össze. [12]

## Milyen programok írásához jó?

A Python egy jól megtervezett programozási nyelv, amely kiválóan alkalmas a valós világban előforduló problémák programozói megoldására. Az alábbiakban összeszedem a Python nyelven írható leggyakoribb programtípusokat:

* rendszer program
  + Az operációs rendszer szolgáltatásaival való együttműködéshez létrehozott belső Python interfészek a Pythont megfelelő nyelvvé teszik a rendszer programok létrehozásához. Ezek az interfészek bizonyos funkciókat biztosítanak, például: fájlok és könyvtárak műveletei, párhuzamos feldolgozás. A rendszererőforrásokkal való együttműködéshez is tartalmaz néhány eszközt, például: környezeti változók, fájlok, socketek, folyamatok, parancssor, szabványos stream-interfészek, shell programozás. [12]
* grafikus felhasználói felület (GUI)
  + A Tkinter egy szabványos objektumorientált felület, amelyet a Python fordító használ. Ez biztosítja a GUI tervezésének alapvető eszközeit. A generált GUI-k pmw formátumban vannak, és hasonló a megjelenésük a különböző platformokon. Egy másik felület a wxPython, amely a C++ könyvtárra épül. A Tkinter és a wxPython az alapvető interfészek a GUI-k tervezéséhez a Pythonban. Ezen kívül számos magas szintű GUI interfész létezik, amelyek megkönnyítik a munkát. [12]
* hálózati és internetes program
  + Különböző modulok vannak beágyazva a Python szabvány könyvtárába, amelyek számos eszközt kínálnak a hálózati programozók számára, például: kliens-szerver kapcsolat, socket programozás, FTP, Telnet, e-mail funkciók, RPC, SOAP. Néhány harmadik féltől származó eszköz, például a mod-Python, lehetővé teszi az apache-hoz hasonló webszerverek számára a Python-szkriptek futtatását. Továbbá néhány népszerű program, például: Django, Turbo gears, Pylon, Zope és Web Ware támogatja a Python szkripteket. [12]
* adatbázis
  + A legtöbb általános adatbázist, mint például a Sybase, az Oracle, az Informix, a MySQL, a PostgreSQL, az SQLite, támogatják a Python interfészek. Az előre definiált adatbázis API-k segítik a programozókat abban, hogy hasonló kódokon keresztül különböző adatbázisokkal dolgozzanak. A Pickle egy szabványos modul, amellyel fájlokban tárolhatjuk és helyreállíthatjuk az objektumokat. Ezen kívül a ZODB egy tisztán objektumorientált eszköz az adatbázisokkal való munkához. A Python 2.5-től kezdődően az SQLite-t a Python szabványos részének tekintik. [12]
* numerikus program
  + A NumPy modullal rendelkező Python hatékony alternatíva lehet a FORTRAN és a C++ számára, mivel hatékony eszközöket kínál matematikai könyvtárakkal való együttműködéshez, egyszerű Python kódok használatával. Az interneten számos harmadik féltől származó eszköz található a Python által végzett numerikus programozáshoz. [12]
* komponens integritás
  + A Python képes integrált kapcsolatot létesíteni kódjai és más alkatrészek között. Az olyan eszközök, mint a Swing és a SIP, importálhatják más nyelvek fordított kódjait a Pythonban történő használatra. [12]

# Gépi látás

A gépi látás lényege, hogy egy modellt alkossunk a valódi világról képek segítségével. A gépi látás rendszere hasznos információkat gyűjt a képről a kétdimenziós vetületei segítségével. Mivel a képek a háromdimenziós világ kétdimenziós vetületei, ezért az információk nem állnak közvetlenül rendelkezésre, vissza kell állítani azokat. Ez a helyreállítás megköveteli a több az egyhez (many-to-one) leképezés inverzióját. Az információk visszaállításához kell a képen szereplő objektumok ismerete, és vetületgeometria is szükséges. [7]

## Alkalmazási területei

Számos terület kapcsolódik a gépi látáshoz. Mint ebből a részből kiderül, sok területen fejlesztettek ki különböző technikákat a képekből való információk kinyerésére. Ebben a részben röviden bemutatok néhány területet, amelyek szorosan kapcsolódnak a gépi látáshoz.

* Képfeldolgozás
  + A képfeldolgozás egy jól fejlett terület. A képfeldolgozási technikák általában képeket alakítanak át más képekké: az információ-helyreállítás a felhasználó feladata. Ez a terület olyan témákat tartalmaz, mint a képjavítás, a képtömörítés és az elmosódott vagy életlen képek javítása. Másrészt a gépi látás algoritmusai képeket vesznek bemenetként, de más típusú kimeneteket hoznak létre, például ábrázolják az objektum kontúrjait egy képen. A gépi látás hangsúlyt fektet az információk automatikus helyreállítására, minimális emberi interakcióval. Általában bizonyos információk javítására és a zaj elnyomására szolgálnak. [7]
* Grafika
  + A számítógépes grafika olyan geometriai primitívekből állít elő képeket, mint a vonalak, körök és szabad formájú felületek. A számítógépes grafikai technikák jelentős szerepet játszanak a megjelenítésben és a virtuális valóságban. A számítógépes grafika a képek szintézise, a gépi látás pedig a képek elemzése. Kezdetben e két terület között nem volt sok kapcsolat, de az utóbbi években ez a két terület egyre szorosabbá vált. A gépi látás görbe és felületi ábrázolásokat, valamint a számítógépes grafikák számos egyéb technikáját használja, a számítógépes grafika pedig a gépi látástól kezdve számos modell felhasználásával valósítja meg a valósághű képeket. A vizualizáció és a virtuális valóság közelebb hozta ezt a két területet. [7]
* Mintafelismerés
  + A mintafelismerés a numerikus és a szimbolikus adatokat osztályozza. Számos statisztikai és szintaktikai technikát fejlesztettek ki a minták osztályozására. A mintafelismerés technikái fontos szerepet játszanak a gépi látásban és az objektumok felismerésében. Valójában számos ipari alkalmazás nagyban támaszkodik a mintafelismerésre. Az objektumfelismerés általában más technikákat igényel. [7]
* Mesterséges intelligencia
  + A mesterséges intelligencia az intelligens rendszerek tervezésével és az intelligencia számítási aspektusainak tanulmányozásával foglalkozik. A mesterséges intelligenciát a képek elemzésére használják a képek tartalmának szimbolikus ábrázolásával, a jellemzőik megszerzése érdekében. A mesterséges intelligenciának három szakaszát tekinthetjük: észlelés, megismerés és cselekvés. Az észlelés a világból érkező jeleket szimbólummá alakítja, a megismerés manipulálja a szimbólumokat, a cselekvés pedig a szimbólumokat olyan jelekké alakítja, amelyek befolyásolják a világ változását. A mesterséges intelligencia számos technikája fontos szerepet játszik a gépi látás minden aspektusában. Valójában a gépi látást gyakran a mesterséges intelligencia részterületének tekintik. [7]
* Neurális hálózatok
  + A neurális hálózati alapú eszközöket gyakran használják az alakzatok jelenlétének meghatározására, vagy arra, hogy a képen lévő objektum jó vagy rossz. Ezek az eszközök a képosztályozóként ismert algoritmusok csoportjába tartoznak, kezdve a példányalapú osztályozóktól, például a k-legközelebbi szomszédtól (k-NN) a döntési fa osztályozókig. Ezek közül számos felhasználható a gépi látásban. Az olyan vállalatok, mint az MVTec Software, már előre képzett ideghálózatokat, támogató vektorgépeket, Gauss-keverék modelleket és k-NN osztályozókat kínálnak. Meg kell jegyezni, hogy a mély tanulási hálózatok több százezer mintaképet igényelhetnek minden hibakategóriához az érvényes felismerési eredmények eléréséhez. Azonban az átfogó mély tanulási hálózatoknak viszonylag kevés mintaképre van szükség a képzéshez, ami jelentős időt takarít meg. [7]

# A program megtervezése

Mint azt korábban már leírtam, a szakdolgozat célja egy zenei kotta digitalizálását végrehajtó szoftver elkészítése. A projekt sikeres megvalósításához a programozás, a gépi látás és képfeldolgozás különböző eszközeit kell jól kezelni. Ezen három terület ismeretével és megfelelő alkalmazásával lehet létrehozni a kívánt programot. A következő szegmensben bemutatom azokat az eszközöket és technikákat, amelyeket fel fogok használni a projekt megvalósítása során.

A korábbi részekben már kifejtettem, hogy a Python programozási nyelvet fogom használni a program megírásához. A Python programnyelvben belül található egy csomag, amit OpenCV-nek hívnak. Az OpenCV a gépi látás, a gépi tanulás és a képfeldolgozás hatalmas, nyílt forráskódú könyvtára. Használatával képeket és videókat lehet feldolgozni a tárgyak, arcok vagy akár kézírás azonosítására. Különböző könyvtárakkal, például a Numpy-al integrálva a Python képes feldolgozni az OpenCV tömbstruktúráját elemzés céljából. A képminták és különféle jellemzőinek azonosításához vektorokat használ, és matematikai műveleteket hajt végre ezeken a tulajdonságokon. [8][11]

Nagyon sok alkalmazás megoldható az OpenCV használatával, ezek közül néhány:

* arcfelismerés
* forgalomszámlálás
* autók sebességének mérése
* hibák észlelése gyártási folyamatokban
* google utcakép (street view)
* videó/kép keresés és visszaállítás
* sofőr nélküli autó navigáció és vezérlés
* objektum felismerés
* orvosi képelemzés

[8][11]

## Thresholding

A gépi látás rendszerek egyik legfontosabb problémája az objektumokat ábrázoló részképek azonosítása. Ez a művelet, amely annyira természetes és olyan egyszerű az emberek számára, meglepően nehéz a számítógépek számára. A kép régiókra osztását szegmentációnak nevezzük. Ideális esetben a partíció egy objektumot vagy annak egy részét képviseli. Formálisan a szegmentálást úgy határozhatjuk meg, hogy egy képet, az F [i, j] -t részképekre, régiókra osztjuk, oly módon, hogy minden részkép egy lehetséges objektum. [7]

A szegmentálás a pixelek régiókba csoportosítása, oly módon hogy:

(1)

(2)

Az (1) egyenletben a {Pi} egy kimerítő particionálás.

A (2) egyenletben a {Pi} egy exkluzív particionálás.

Minden Pi régió kielégít egy alapot, vagyis a partició minden pontjának van valamilyen közös tulajdonsága. A szomszédos régiókhoz tartozó pixelek, ha együtt veszik őket, nem felelnek meg az alapnak. [7]

Amint a fentiekből látható, a partició megfelel az alapnak. Ez az alap lehet rendkívül egyszerű is, például egyenletes intenzitású, de a legtöbb alkalmazásban összetettebb. A szegmentálás nagyon fontos lépés a képek megértésében. [7]

A szürkeárnyalatos kép megfelelő szegmentálásával bináris képet kapunk. Ha egy objektum intenzitási értékei intervallumban vannak, és a háttérpontok intenzitási értékei ezen az intervallumon kívül esnek, akkor bináris képet kaphatunk egy thresholding művelet segítségével, amely az intervallum pontjait 1-re, a tartományon kívül pedig 0-ra állítja. Így a bináris látás esetében a szegmentálás és a threshold azonos. Sok kamerát terveztek a thresholding művelet hardveres végrehajtására. Az ilyen kamera kimenete bináris kép. A legtöbb alkalmazásban azonban a kamerák szürkeárnyalatos képet adnak, és a bináris képet threshold alkalmazásával nyerik ki. [7]

A thresholding egy olyan módszer, amellyel a szürkeárnyalatos képet bináris képpé alakíthatjuk át úgy, hogy az érdeklődésre számot tartó objektumok elkülönüljenek a háttértől. Ahhoz, hogy a thresholding hatékony legyen az objektum és a háttér szétválasztásban, szükséges, hogy az objektumok és a háttér közötti kontraszt kellően nagy legyen, és ismerjük az objektumok vagy a háttér intenzitásának szintjét. Fix thresholding sémában ezek az intenzitási jellemzők határozzák meg a thresholdot. [7]



1. ábra: A kép bal oldalán a szürkeárnyalatos kép, jobb oldalán a thresholdolt kép látható.

Forrás: https://pippin.gimp.org/image-processing/gen\_images/threshold.png (megtekintve: 2021.04.25.)

Tegyük fel, hogy a B [i, j] bináris kép megegyezik az FT [i, j] thresholdolt szürke képpel, amelyet az eredeti F [i, j] szürke kép T thresholdjának felhasználásával kapunk. Így,

(3)

sötétebb objektum világos háttéren:

(4)

Ha tudjuk, hogy az objektum intenzitási szintje [T1, T2] intervallumba esik, akkor használhatjuk:

(5)

Az általános thresholding sémát, amelyben egy objektum intenzitási szintje több diszjunkt intervallumból származhat, ábrázolhatjuk úgy, hogy:

(6)

ahol Z az objektumkomponensek intenzitási értékeinek halmaza. [7]

Vegyük figyelembe, hogy az alkalmazás tartományról szóló tudás hogyan épül be a thresholding algoritmusba. Valójában a tartományra szabták; ezért ugyanazok a thresholdok nem biztos, hogy működnek egy új tartományban. A thresholdot általában az alkalmazás tartományával kapcsolatos tapasztalatok alapján választják ki. Bizonyos esetekben a rendszer első néhány futtatását fel lehet használni egy kép interaktív elemzésére és a threshold megfelelő értékének meghatározására. [7]

A képek automatikus thresholdja gyakran az első lépés a képek elemzésében a gépi látás rendszerekben. Számos technikát fejlesztettek ki a kép intenzitáseloszlásának és az objektumok felismerésének felhasználására, a megfelelő threshold automatikus kiválasztására. [7]

## Vonalkeresés

A zenei kották meglehetősen sok egyenes vonalat tartalmaznak. Az öt hosszú fő vonalat kottavonalaknak hívják, ezek mindig egyenesek és párhuzamosak. Ezeken kívül az egyes hangjegyek is tartalmaznak vízszintes, illetve függőleges vonalakat is. Ezeket a vonalakat fel kell ismerni és meg kell határozni az elhelyezkedésüket.

### Hough Line Transform

A Standard Hough Transform (továbbiakban SHT) egy körök, vonalak és egyéb egyszerű alakzatok megtalálására alkalmas módszer. Az eredeti Hough Transform csak vonal transzformáció volt, amely viszonylag gyors megoldás egy vonal megtalálására egy bináris képen. [1][3][7]

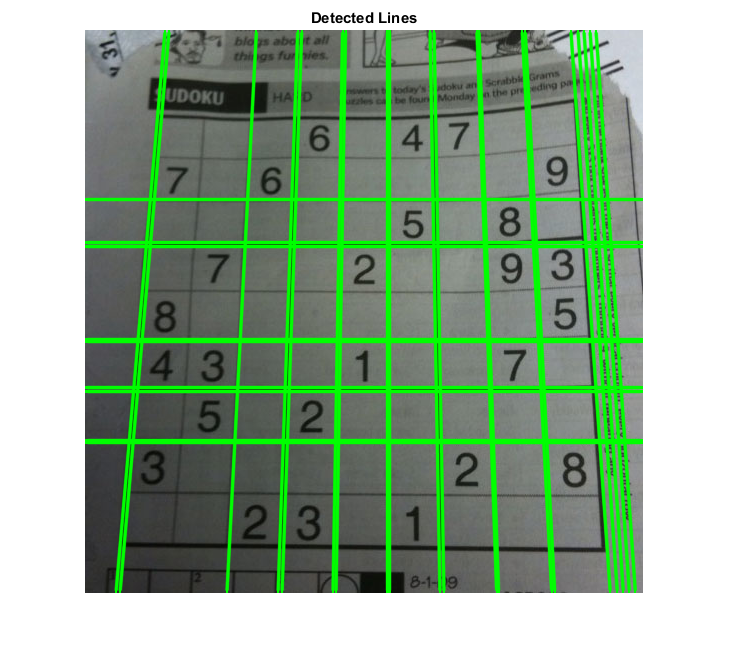
Az SHT alapelmélete az, hogy a bináris kép bármely pontja része lehetne a lehetséges vonalaknak. Ha az egyes vonalakat ’a’ meredekség és ’b’ metszéspont alapján paraméterezzük, akkor az eredeti kép egy pontja átalakul az (a, b) síkbeli pontok azon helyévé, amely megfelel az adott ponton áthaladó összes vonalnak. Ha a bemeneti kép minden pixeljét, ami nem nulla, a kimeneti kép ilyen pontjainak halmazává konvertáljuk, és összesítjük az összes ilyen átalakítást, akkor a bemeneti képen megjelenő vonalak helyi maximumokként jelennek meg a kimeneti képen. Mivel az egyes pontokból származó átalakításokat összegezzük, az (a, b) síkot általában felhalmozó síknak nevezzük. [1][3][7]

Előfordulhat, hogy a meredekség és metszéspont módszere nem igazán a legjobb módja annak, hogy az összes pontot áthaladó vonalat ábrázolja, mivel a meredekség függvényében a vonalak jelentősen eltérő gyakoriságúak, valamint a lehetséges meredekségek intervalluma végtelen lehet. Éppen ezért némileg eltér a numerikus összeállításban használt transzformációs kép tényleges paraméterezése. Az előnyben részesített paraméterezés mindegyik egyenest poláris koordinátákban (p, 0) jeleníti meg, az implicit vonal a jelzett ponton áthaladó, de az origótól az adott pontig sugárirányra merőleges vonal. Egy ilyen egyenes egyenlete:

(7)

Az (x, y) az egyenes vonalon lévő pontok koordinátáit, vagyis a pixelteret jelöli, míg P és 0 a paramétertér. Az összes (P, ()) paraméter esetében az algoritmus meghatározza, hogy van-e elegendő bizonyíték egy valódi vonalra az adott pixelnél, amely kiszámítja a paraméterértékeket és felhalmozza a tárolókat a paramétertérben az összes képpont alapján. [1][3][7]

A Hough transzformációs számítás három részből áll: (1) a paraméterértékek kiszámítása és a tárolók felhalmozása a paramétertérben; (2) a vonalszakaszokat ábrázoló lokális maximumok megkeresése; és (3) a vonalszakaszok kivonása az ismeretek felhasználásával a maximális pozíciókból. [1][3][7]



2. ábra: A Standard Hough Line Transform által felismert vonalak egy Sudokun

Forrás: http://amroamroamro.github.io/mexopencv/opencv/hough\_lines\_demo\_02.png (megtekintve: 2021.04.26.)

Az algoritmus egyszer végigmegy a kép minden pixelén. Ezért időbeli bonyolultsága lineáris a teljes képszámmal, amely a kép szélességének és magasságának szorzata. Ugyanakkor, mivel az algoritmusnak fel kell halmoznia az egyes pixelek értékeit, tárolási igénye a pixelszámmal együtt változik. Tehát az algoritmus számítási komplexitása és tárolási követelményei jelentik az SHT fő szűk keresztmetszeteit. Ezért az SHT számos változatát javasolták a számítási és tárolási terhek enyhítésére. [1][3][7]

Az OpenCV Hough transzformációs algoritmus nem teszi egyértelművé ezt a számítást a felhasználó számára. Ehelyett egyszerűen visszaadja a helyi maximumokat a (p, 0) síkban. Azonban a vonaltranszformációs függvény argumentumainak megértéséhez azonban ezt a folyamatot meg kell értenünk. [1][3][7]

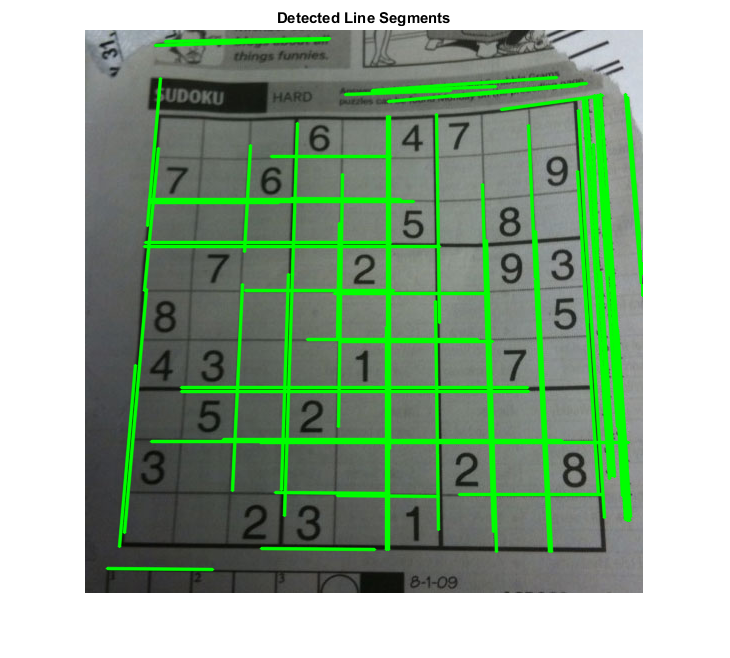
### Probabilistic Hough Transform

A Probabilistic Hough Transform (továbbiakban PHT) ennek az algoritmusnak egy variációja, amely többek között az elhelyezkedése mellett, a méretét is kiszámítja a vonalaknak. Azért ’probabilistic’ (valószínűségi), mert ahelyett, hogy felhalmozná az összes lehetséges pontot az felhalmozás síkjában, csak a töredéküket halmozza fel. Az elképzelés az, hogy ha a keresett dolog mindenképp elég nagy lesz, akkor az adott idő töredéke is elég lesz, hogy megtaláljuk; ennek a valószínűségi módszernek az eredménye jelentősen csökkentheti a számítási időt. Mindkét algoritmus ugyanazzal az OpenCV függvénnyel érhető el, bár néhány argumentum jelentése eltérő lehet attól függően, hogy melyik módszert használjuk. [3][5]

Az első argumentum a bemeneti kép. 8 bites képnek kell lennie, de a bemenetet bináris információként kezeljük. A második argumentum egy olyan hely mutatója, ahol az eredmények tárolhatók, amely lehet memóriatároló vagy sima [N,1] mátrix tömb. A következő argumentum a metódus, ami lehet CV\_HOUGH\_STANDARD, CV\_HOUGH\_PROBABILISTIC vagy CV\_HOUGH\_MULTI\_SCALE a PHT-hoz, vagy a SHT többskálás változatához. [3][5]

A következő két argumentum az rho és theta. Ezek adják meg a vonalak kívánt felbontását (a felhalmozó sík felbontását). Az rho egységei a pixelek, a téta egységei a radiánok; így az felhalmozó sík egy kétdimenziós hisztogramnak tekinthető, ahol a két dimeznió cellái az rho pixelei és theta radiánjai. A threshold a felhalmozó síkban az az érték, amelyet el kell érni ahhoz, hogy a rutin felismerje a vonalat. Ez az utolsó argumentum a gyakorlatban kissé trükkös, ugyanis nincs normalizálva, ezért számíthatunk arra, hogy a kép méretét fel kell nagyítanunk. Azt se feledjük, hogy ez az argumentum valójában azon a pontok számát is megadja, amelyeknek az adott vonalon kell lennie, hogy a rutin felismerje a vonalat. [3][5]

A param1 és param2 argumentumokat az SHT nem használja. A PHT esetében a param1 meghatározza a felismerendő vonalszakasz minimális hosszát, a param2 pedig az egy egyenesre eső szegmensek közötti elválasztást, amely szükséges ahhoz, hogy az algoritmus ne egyesítse őket egyetlen hosszabb szakaszba. A többskálás HT esetében a két paraméter arra szolgál, hogy nagyobb felbontást érjünk el, amelyekre a vonalak paramétereit ki kell számítani. A többskálás HT először kiszámítja a vonalak helyét az rho és a theta paraméterek által megadott pontossággal, majd tovább finomítja ezeket az eredményeket param1 és param2 tényezőkkel. [3][5]



3. ábra: A Probabilistic Hough Transform által felismert vonalak egy Sudokun

Forrás: http://amroamroamro.github.io/mexopencv/opencv/hough\_lines\_demo\_03.png (megtekintve: 2021.04.26.)

A függvény eredménye attól függ, hogyan hívták meg. Ha a line\_storage értéke egy mátrix tömb volt, akkor a tényleges visszatérési érték NULL lesz. Ebben az esetben a mátrixnak CV\_32FC2 típusúnak kell lennie, ha az SHT-t vagy a többskálás HT-t használja, és CV\_32SC4-nek kell lennie, ha a PHT-t használja. Az első két esetben az egyes sorok p és 0 értékei a tömb két csatornájába kerülnek. A PHT esetében a négy csatorna megtartja a felismert szegmensek kezdő és végpontjának x és y értékeit. Mindegyik esetben a tömb sorainak számát a cvHoughLines2 () frissíti, hogy megfelelően tükrözze a megtalált vonalak számát. [3][5]

Ha a line\_storage érték egy memória tároló mutatója volt, akkor a visszatérési érték egy CvSeq szekvencia struktúra mutatója lesz. Ebben az esetben az egyes vonalakat vagy vonalszegmenseket az alábbi paranccsal kapjuk meg:

float\* line = (float\*) cvGetSeqElem(lines , i);

ahol a lines a cvHoughLines2() visszatérési értéke, az i pedig a keresett vonal indexe. Ebben az esetben a vonal egy mutató lesz az adott vonal adataihoz, a line[0] és a line[1] pedig a lebegőpontos p és 0 értékek lesznek, vagy a CvPoint struktúrák lesznek a szegmensek végpontjaihoz. [3][5]

## Kontúrkeresés

A kontúrokat egyszerűen úgy lehet megmagyarázni, hogy egy objektum határai mentén az összes azonos színű vagy intenzitású folytonos pontot összekötő görbe. A kontúrok hasznos eszközök az alakzatok elemzéséhez, valamint az objektumok felismeréséhez. A pontosabb eredmény érdekében bináris képeket kell használni, tehát mielőtt kontúrokat keresnénk, mindenképpen thresholding kell, ha eddig nem történt volna meg. Az OpenCV-ben a kontúrok megkeresése során, fehér tárgyat keresünk fekete háttéren. Tehát a keresett objektumnak fehérnek, a háttérnek pedig fekete színűnek kell lennie. [10][11]

Az OpenCV-ben a findContours() függvényt alkalmazhatjuk a kontúrok megkeresésére. Három argumentum van a cv2.findContours() függvényben, az első az forrás kép, a második a kontúr visszakeresési módja, a harmadik a kontúr közelítési módszere. És ez adja ki a képet, a kontúrokat és a hierarchiát. A CONTOURS a kép összes kontúrjának listája. Minden egyes kontúr egy Numpy tömb, amiben (x, y) az objektum határpontjainak koordinátái. [10][11]

Általában a cv2.findContours() függvényt használjuk a képen lévő objektumok felderítésére. Néha az objektumok különböző helyeken vannak, de bizonyos esetekben egyes alakzatok más alakzatokon belül vannak. Ebben az esetben a külsőt szülőnek, a belsőt pedig gyermeknek nevezzük. Így a kép kontúrjai valamilyen kapcsolatban állnak egymással. És megadhatjuk, hogy az egyik kontúr hogyan kapcsolódik a másikhoz, például, hogy valamilyen más kontúr gyermeke-e, vagy szülője-e. Ennek a kapcsolatnak az ábrázolását Hierarchiának nevezzük. [10][11]

A visszakeresési módoknak több fajtája is van, ezek a következőek:

* cv2.RETR\_LIST
  + Ez a négy mód közül a legegyszerűbb. Egyszerűen lekéri az összes körvonalat, de nem hoz létre szülő-gyermek kapcsolatot. A szülők és a gyerekek ebben a szabályban egyenlőek, és csak sima kontúrok, azaz mindegyik azonos hierarchia szinthez tartozik. [10][11]
* cv2.RETR\_TREE
  + Ez az módszer lekéri az összes körvonalat, és létrehoz egy teljes családi hierarchia listát. Még azt is megmondja, ki a nagypapa, apa, fiú, unoka és még azon túl is. [10][11]
* cv2.RETR\_CCOMP
  + Ez a mód lekéri az összes kontúrt és 2-es szinttű hierarchiába rendezi őket, vagyis az objektum külső kontúrjai az 1-es hierarchiába kerülnek. Az objektum belsejében lévő lyukak vagy egyéb alakzatok kontúrjai (ha vannak) a 2-es hierarchiába kerülnek. [10][11]
* cv2.RETR\_EXTERNAL
  + Ha ezt a módot használjuk, akkor csak a legkülső kontúrokat adja vissza. Minden gyermek kontúr elmarad. [10][11]

A cv2.findContours() harmadik argumentuma a közelítési mód. Fentebb említettem, hogy a kontúrok az azonos intenzitású vagy színű alakok határai. Az alakzat határának (x, y) koordinátáit tárolja. De tárolja-e az összes koordinátát? Ezt a kontúr közelítési módszere határozza meg. [10][11]

Két különböző mód létezik, az első a cv2.CHAIN\_APPROX\_NONE, amely a megtalált elem összes határpontját eltárolja. Ez esetenként azonban felesleges, hiszen ha egy vonalat keresünk akkor nincs szükség az összes pontjának eltárolására. Elég számunkra, ha a végpontjait meghatározzuk. Ezt a cv2.CHAIN\_APPROX\_SIMPLE módszerrel tudjuk megvalósítani, amely eltávolítja az összes felesleges pontot és tömöríti a kontúrt, ezáltal memóriát takarít meg. [10][11]

A kontúrok megrajzolásához a cv2.drawContours() függvényt használjuk. Bármilyen alakzat megrajzolására használható, feltéve, hogy megvannak a határpontjai. Első argumentuma a forráskép, a második argumentuma a kontúrok, amelyeket listaként kell neki átadni, a harmadik argumentum a kontúrindex, a többi argumentum pedig a szín, a vastagság és egyéb állítható paraméterek. [11]



4. ábra: Példa a kontúrok keresésére és megrajzolására OpenCV használatával

Forrás: https://robottini.altervista.org/wp-content/uploads/2012/05/contours.png (megtekintve: 2021.04.27.)

## Konvolúciós neurális hálózat

Az objektum felismerés célja a célobjektumok detektálása a képfeldolgozás és a mintafelismerés elemeivel és módszereivel, valamint az objektumok szemantikai kategóriáinak meghatározása és a célobjektum konkrét helyzetének megjelölése a képen. Gyakorlati alkalmazásban nagy kihívást jelent a számítógépes technológia használata az objektumok automatikus felismerésére. A komplex háttér, a zavaró zajok, az alacsony felbontás, a méretek és a pozíciók változásai is mind komolyan befolyásolják az objektumok észlelését. A hagyományos módszer egy kézzel készített funkción alapul, amely egyes nehézségeket nem kezel jól, és általános használatra sem a legalkalmasabb. Emiatt számos módszert dolgoztak ki az objektum észlelés teljesítményének javítására. [4][6][13]

A kovolúciós neurális hálózat (Convolutional Neural Network) egy olyan mély tanulási modell, amely képes a hierarchikus tanulásra, valamint kutatások bizonyítják, hogy az általa használt funkciók jobban alkalmazhatóak az objektumokok szegmentálására és felismerésére, mint a kézzel készített verzió. [4][6][13]

A gépi látás területén a jellemzők kinyerése és osztályozása mindig is nagyon fontos kutatási irány volt. A hagyományos képfeldolgozási feladatokban a kinyert jellemzők gyakran előre megtervezett jellemzők, statisztikai törvényszerűségek vagy előzetes ismeretek alapján történtek. Tehát nem képes az eredeti kép információit átfogóan és pontosan ábrázolni. [4][6][13]

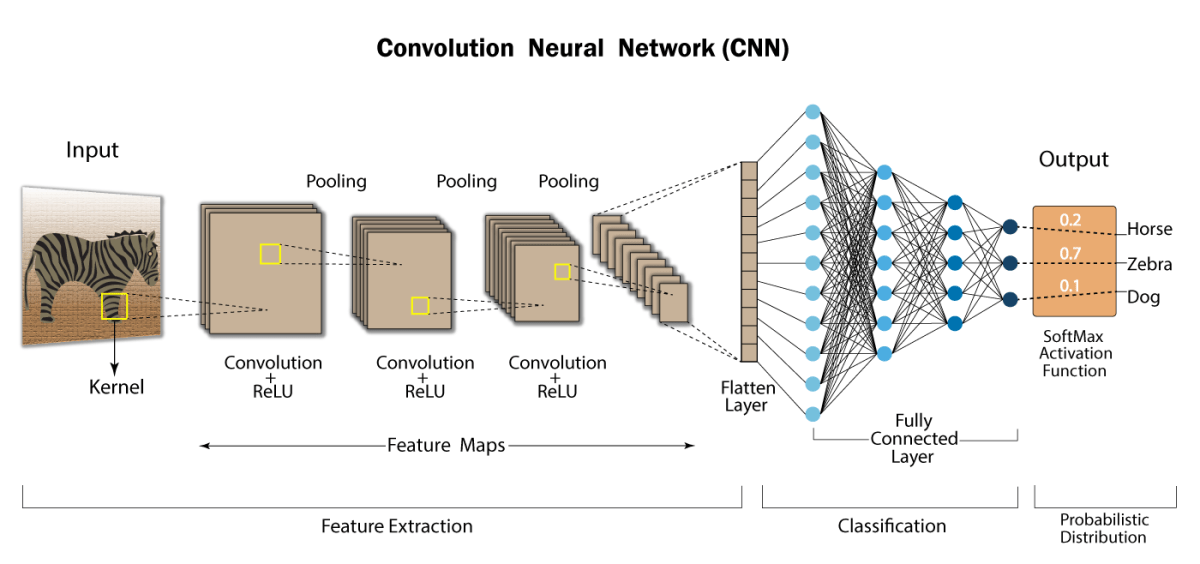
A Convolutional Neural Network (továbbiakban CNN) egy végpontok közötti tanulási modellt nyújt, amelyben a paraméterek a gradiens ereszkedési módszerrel oktathatók. A jól képzett CNN teljesebben elsajátíthatja a kép jellemzőit, és úgy tekinthetünk rá, mint egy fejlettebb fekete dobozra, amely ki is bontja a jellemzőket. [4][6][13]

Az ideghálózat fontos ágaként a CNN növeli a befogadó mező és a megosztott súly fogalmát, ami nemcsak nagymértékben csökkenti a tanítás paramétereit, hanem csökkenti a hálózati modell komplexitását is. Az egyes rétegek jellemzői az előző réteg helyi területéből (befogadó mezőből) származnak a konvolúciós kernel súlyának megosztásával. Ezek a jellemzők teszik a CNN-t alkalmassá a kép jellemzőinek megtanulására és reprezentálására. [4][6][13]

Egy tipikus CNN-ben az első néhány réteg általában váltakozó konvolúciós és összevonási réteg, és a kimeneti réteg közelében lévő hálózat utolsó rétegei általában egy teljesen összekapcsolt hálózat. A CNN oktatása főként az előterjesztő (Forward Propagation) és a visszaterjesztő (Back Propagation) algoritmust használja a rétegkapcsolati súlyok, torzítás és egyéb paraméterek megismerésére. Az oktatás egy felügyelt tanulási folyamat, amely megköveteli a képadatok bevitelét és a megfelelő címkéket a hálózati paraméterek optimalizálásához, melyek alapján végül egy optimalizált súlyú modellt kapunk. [4][6][13]

A CNN különböző funkcionális rétegstruktúrából áll. A tipikus CNN rendelkezik konvolúciós réteggel, összevonási réteggel és teljesen összekapcsolt réteggel. A CNN azonban számos új réteget adhat hozzá az evolúció és a fejlesztés folyamatához, például az SPP (Spatial Pyramid Pooling) réteget, a ROI (Region of Interest) egyesítő rétegét és a R-CNN (Region Based Convolusional Neural Network) réteget. A hagyományos CNN szerkezet javításával jobb teljesítmény érhető el a specifikus problémák megoldására. [4][6][13]

### A hagyományos CNN szerkezete



5. ábra: A hagyományos CNN rétegei

Forrás: https://i0.wp.com/developersbreach.com/wp-content/uploads/2020/08/cnn\_banner.png?fit=1200%2C564&ssl=1 (megtekintve: 2021.04.27.)

Amint az ábrán látható, a tipikus CNN struktúra főként a bemeneti rétegből, a konvolúciós rétegből, az összevonási rétegből, a teljesen összekapcsolt rétegből és a kimeneti rétegből áll. A konvolúciós ideghálózat bemenete általában az eredeti kép. az l-edik réteg j-edik jellemzőtérképét jelöli. Feltételezzük, hogy egy jellemzőtérkép a konvolúciós rétegen, és az alábbiak szerint generálódik:

(8)

A az l réteg funkció térképének megfelelő súlymátrix, a \* a konvolúció szimbóluma, amely specifikus szűrőket ad az l-1 funkció térképeihez, majd a számított eredményeket hozzáadjuk a eltéréshez, végül az –t az f(x) függvény nem lineáris aktiváló műveleteivel kapjuk meg. [13]

Az összevonási réteg, más néven lemásoló réteg, általában a konvolúciós réteget követi, és lemásolja az előző jellemzőtérképet adott szabályok szerint. Ezek a szabályok a következőek: maximum összevonás, átlagos összevonás, sztochasztikus összevonás és átfedő összevonás. Az összevonási rétegnek két fő aspektusa van: a funkció térképek dimenzióinak csökkentése, és a skálák változatlanul tartása. Feltételezzük, hogy egy jellemzőtérkép az összevonási rétegben, és az összevonási műveletet az alábbi formula írja le:

(9)

A pooling(x) függvény a lemásolási szabályt jelöli, pedig az összevonás súlya. Általában egy fix érték, a b eltérést és az f(x) aktiváló műveletet pedig nem használjuk. Tehát az összevonási műveletek általános formája:

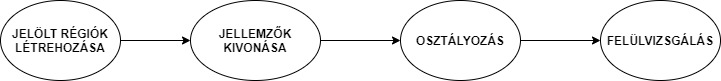
(10)

Teljesen összekapcsolt hálózatban a képek jellemzőtérképeit egydimenziós jellemzővektorokká fűzzük, majd bemeneteként odaadjuk a teljesen összekapcsolt hálózatnak. A teljesen összekapcsolt réteg kimenetét úgy kapjuk meg, hogy a bemenethez súlyozott összegzéseket adunk, majd az aktiváló funkciót alkalmazzuk az alábbi módon:

(11)

[13]

### Objektum felismerés a CNN használatával



6. ábra: A tradícionális objektum felismerés lépései (saját szerkesztés)

Amint az ábrán látható, a hagyományos objektumdetektáló rendszer általában négy szakaszra oszlik. Először is, az adott képen jelölt régiókat hozunk létre csúszóablakos megoldással, majd a releváns jellemzőket kivonjuk ezekből a régiókból. Ezután a betanított osztályozóval a régiók osztályozására és azonosítására kerül sor, végül az eredmények felülvizsgálata és optimalizálása történik meg. [13]

* A jelölt régiók létrehozása

Ez a szakasz az objektum helyének megszerzésére szolgál. Az objektum azonban bárhol lehet a képen. Az objektum mérete és helyzete is bizonytalan, ezért különböző méret- és képarányú csúszóablakokkal végigjárjuk az egész képet. Ez a stratégia szinte az összes helyet magába foglalja, ahol az objektum elhelyezkedhet, de hátrányai is vannak, például az időigény és a redundáns ablakok. Ez súlyosan befolyásolhatja a jellemzők kivonásának gyorsaságát és teljesítményét. [13]

* Jellemzők kivonása

A jellemzők kivonása közvetlenül befolyásolja az osztályozó tervezését és teljesítményét. Nehéz azonban megfelelő funkciót megtervezni a különféle külső tényezők, például a tárgyak mozgása, a megvilágítás változása, az összetett és megváltoztatható környezet miatt. Néhány kézzel készített funkciót széles körben használnak ebben a szakaszban, mint például a skála változatlan jellemzőinek átalakítása (Scale Invariant Feature Transform), az orientált színátmenetek hisztogramjai (Histograms of Oriented Gragients) és a helyi bináris minták (Local Binary Patterns). [13]

* Osztályozás

Az SVM vagy az AdaBoost osztályozókat széles körben használják a kivont jellemzők osztályozására ebben a szakaszban. [13]

* Felülvizsgálás

A besorolás után még mindig sok redundáns ablak van, ezért el kell távolítani a redundáns ablakokat, és optimalizálni kell az észlelési eredményeket a NMS (Non-Maximum Suppression) és az átfedésben lévő határoló doboz (Bounding Box) kombinálásával. A hagyományos objektumészlelésnek két fő problémája van: egyrészt a csúszó ablakon alapuló módszer nem elég céltudatos, nagy időbonyolultságú, és túl sok redundáns ablak van; másodszor, a kézzel készített funkciók nem elég robusztusak a változások sokféleségéhez. [13]

## Régiókon alapuló konvolúciós neurális hálózat

A redundáns ablakok problémájának megoldására a régiójavaslat jó megoldást kínál. Előre kitalálja a képen található obejktumok lehetséges helyét. A régiójavaslat jól használja ki a kép információit, például a színt, az éleket és a textúrát, ami biztosíthatja, hogy a rendszer viszonylag nagy visszahívási lehetőséget kapjon kevesebb ablak esetén. Ez nagymértékben csökkenti a következő műveletek időigényét, és a régiójavaslatok minősége magasabb, mint a csúszóablakoké. 2014-ben Ross B. Girshick javaslata alapján a CNN hagyományos csúszóablakos megoldását leváltották a régiójavaslatos CNN objektum felismerési rendszerére. Ez nagy sikert aratott az objektum detektálásban, és megnyitotta a kapukat a hatékonyabb CNN alapú objektum kereséshez. [13]

A Region Based Convolution Neural Network (továbbiakban R-CNN) három modullal rendelkezik: először létrehoz kategóriafüggetlen régiójavaslatokat. A második modul egy nagy CNN, amely kivonja az adott hosszúságú jellemzővektorokat a régiókból. Az utolsó modul szétválasztja az objektumot és a hátteret lineáris SVM (Support Vector Machine) segítségével. Annak érdekében, hogy növeljük a helymeghatározás pontosságát egy lineáris regressziós modellt képeztek ki a detektálódobozok koordinátáinak módosítására. [13]

Azonban ennek a módszernek is megvannak a hátrányai. Az R-CNN túl sok jelölt ablakot generál, és mivel a CNN-en keresztül vonja ki a jellemzőket, ennek az időigénye még mindig túl nagy. A másik probléma az, hogy a CNN-ek egy fix méretű képet kapnak bemenetként, amelyet az R-CNN vágással vagy egyéb művelettel tud a saját méretéhez illeszteni, ami a képarány és a méretarány módosulását is okozza. [13]

He Kaiming kidolgozta az SSP-Net módszert az R-CNN hibáinak kiküszöbölésére. Az SPP-Net csak egyszer futtatja le a konvolúciós réteget a teljes képen, hogy megkapja a jellemzőtérképeket. A számítási feladatok megosztása miatt a lefutási idő 24-64-szer gyorsabb, mint a sima R-CNN-nél. A javasolt régiók méretei eltérőek, de a teljesen összekapcsolt réteg egy fix méretet vár bemenetként. Ennek megoldására jött létre az SPP réteg. Az SPP réteg az utolsó konvolúciós réteg mögé került, a jelölt ablakok szelektív kereséssel kapott pozíciója, az eredeti képhez hasonlóan az utolsó konvolúciós réteg jellemzőtérképeihez vannak rendelve. Ezután a javított jellemzőket a többszintű SPP összevonja, és minden ablakhoz létrehoz egy adott hosszúságú jellemzővektort. Az SPP-Net képes megőrizni mind a globális, mind a helyi információit a képnek, szóval nagyobb pontossággal rendelkezik, mint az R-CNN. [13]

A Fast R-CNN az SPP réteg alapján létrehozta a Region of Interest (RoI) összevonási réteget, amely egy egyszintű SPP rétegnek tekinthető és képes különböző méretű képeket is feldolgozni. A Fast R-CNN módszer az SPP-Nethez képest a következő előnyökkel rendelkezik:

* A Fast R-CNN egy végponttól végpontig tartó objektum felismerési rendszert tanít, az SPP pedig egy többlépéses folyamatot.
* A Fast R-CNN egyszerre tanítja az osztályozást és a határoló dobozok létrehozását.
* Az SPP-net nem tudja frissíteni a konvolúciós rétegeket tanítás közben, a Fast R-CNN azonban képes frissíteni az összes hálózati réteget. Az objektumdetektálási feladathoz a konvolúciós rétegek paramétereinek finomhangolása is szükséges és fontos, mert a megelőző konvolúciós rétegek több helyinformációt tudnak megőrizni, mint a későbbiek. [13]

A régiójavaslatok kiszámítása még így is egy bökkenőt jelent a Fast R-CNN szempontjából. A minőségi régiójavaslatok nagyban befolyásolják az objektum felismerés gyorsaságát és pontosságát. Ennek megoldására találták ki a régiójavaslati hálózatot (Region Proposal Network), amely szinte valós időben képes észlelni az objektumokat. A régiójavaslati hálózat egy teljesen konvolúciós hálózat (Fully Convolutional Network) amely egyidőben jósolja meg az objektumok határait és az objektivitási pontjukat minden lehetséges helyen. Az alapötlete az, hogy a CNN által magas visszahívású régiójavaslatokat generál, amely nemcsak felgyorsítja a felismerési sebességet, hanem a kombinálni lehet a CNN végpont alapú tanításával. [13]

A Faster R-CNN egyesítette a régiójavaslatot és a CNN osztályozását egy végpontok közötti detektálási hálózathba, javítva ezzel mind a sebességet, mind a pontosságot. A Faster R-CNN azonban még mindig nem képes nagy pontossággal teljesíteni a valós idejű objektum-észlelés követelményeit, mivel a CNN jellemzői kivonásának kiszámítása még mindig időigényes. [13]

Összességében az R-CNN, az SPP-Net, a Fast R-CNN és a Fater R-CNN mind a CNN-en alapuló objektum felismerést fejlesztették tovább, hogy gyorsabb és pontosabb módszert kapjanak. Az R-CNN módszerek jelenleg is az objektum felismerés egyik legfőbb ágát képviselik. [13]

# A program megvalósítása

Az elkészítendő program részletes megtervezése után nincs más hátra, mint implementálni a korábban leírtakat. A megvalósításhoz a programozási nyelven kívül (amelyet már korábban részleteztem) szükség van még egy fejlesztői környezetre is. Az én választásom a Visual Studio Code-ra esett, amely egy ingyenes, nyílt forráskódú fejlesztői környezet, mely számos operációs rendszeren elérhető és beépített Git támogatással is rendelkezik. Ezen kívül pedig rengeteg programozási nyelvet támogat, tehát nagyszerűen használható például Python, C++, Java, C#, PHP, SQL és egyéb programok írására.

## Előkészületek

A projekt első lépéseként két függvénykönyvtárat kell letöltenem, majd beimportálnom ahhoz, hogy használni tudjam a funkcióikat. Ezek nem mások, mint az OpenCV és a NumPy. Ezután következik a feldolgozandó kép beolvasása. Ehhez a művelethez először szükségem van néhány képre, amelyeken a tesztelést el fogom végezni. Az interneten számos zenei kotta érhető el jó minőségben, de talán a legnagyobb gyűjtemény a ChoralWiki oldalon található meg, ahol több ezer zenei kotta található meg rendszerezett formában, és tölthető le PDF, Midi, MusicXML és egyéb formátumokban. Mivel számomra valamilyen képformátumú bemeneti adatra van szükségem, így PDF formátumban töltöttem le őket, majd egy internetes konverter segítségével JPG formátumra alakítottam őket. Így már felhasználhatók bemenetként a digitalizáló programhoz.

## Threshold

A könyvtárak importálása és a kép beolvasása után a threshold megadása következik. A beolvasott képet először szürkeárnyalatos képpé kell átalakítani. Ez egyszerűen elvégezhető, hiszen az OpenCV-ben van erre egy beépített metódus, ami az RGB képeket szürkeárnyalatossá teszi. Miután ez elkészült egy bináris képet kell létrehozni ebből a képből, amelyet thresholdolással kapok meg.

## Kottavonalak megkeresése

A zenei kottákban fellelhető hosszú vonalakat, amelyeken a hangjegyek is elhelyezkednek, kottavonalaknak hívjuk. A kottaírásban jellemzően őt kottavonal alkot egy sort. A vonalak bizonyos hangmagasságokat jelölnek, és a hangokat jelölő hangjegyek pedig ezeken a vonalakon, vagy a vonalak között helyezkedhetnek el.

Ahhoz, hogy a program megtalálja és felismerje ezeket a kottavonalakat, a Probabilistic Hough Transformot alkalmazom. Az ehhez szükséges eljárás a cv2.HoughLinesP(), amelynek miután megadtuk a szükséges paramétereit, azaz a képet, amelyen a keresést végzi, a rho-t, a theta-t, a minLineLenght-et és a maxLineGap-et. A korábbiakban ismertetett módon ez az algoritmus elkezdi a képen található adott hosszúságú és töredezésmentes vonalakat megkeresni és felismerni. Ahhoz, hogy jól működjön az eljárás, a thresholdolt képen kell lefuttatni. Miután elkészült a vonalak megkeresésével, a vizuálisabb ellenőrzés érdekében kirajzoltatom a kottavonalakat, amelyeket a HoughLines megtalált.



7. ábra: Vonalfelismerés eredménye

Ezeket a vonalakat eltárolom egy adott listába, hogy a későbbiekben felhasználhatóak legyenek. Ezt a listát egyszerűen csak Lines-nak neveztem el, hogy minél egyértelműbb legyen. Azonban az ebben eltárolt vonalak nem mindig feleltek meg az elvárásoknak, esetenként nem talált meg vonalakat, vagy több vonalat talált meg, mint kellene.

Ezen problémák kiküszöbölésére hoztam létre még három listát, amelyek segítségével javítok a teljes vonalkeresési procedúra sikerességén. Először is y irányú, azaz vízszintes hosszúság szerint sorba rendezem a detektált vonalakat, ás azokat ilyen formában lementem egy goodLines listába. Ezt követően megvizsgálom, hogy van-e szinte azonos pixelhelyen több vonal is, ha akad ilyen, azokat összevonom, hiszen azok nagy valószínűséggel egy és ugyanaz vonal, csak a vastagsága miatt az algoritmus több vonalnak értelmezte (például egy 5 pixel vastag vonalat érzékelhet 5 darab 1 pixeles vonalnak). Ezt a gyakorlatban úgy valósítottam meg, hogy a goodLines listából egyesével hozzáadtam a vonalakat egy tempLines listába, amelyben megvizsgáltam az adott kottavonal y koordinátáját, majd minden esetben összevetem a következő vonal koordinátájával és ha a kettő közötti különbség kevesebb, mint adott pixel, akkor összevontam a két vonalat.

Amikor ezzel elkészültem, a vonalakat a tempLines listából hozzáadtam a permaLines listához, majd mivel y koordináta szerint vannak rendezve, ötösével eltárolom őket, így a kottasorok egyben vannak eltárolva.

## Kottavonalak kitörlése

Ahhoz, hogy a hangjegyeket könnyebben fel tudjam ismerni, a korábban megtalált és megjelölt vonalakat eltávolítom a képről. Így megelőzöm azokat a problémákat, amelyekkel akkor szembesülnék, ha úgy akarnám felismerni a hangjegyeket, hogy a háttérben a vonalak bezavarnának a felismerési algoritmusnak. A vonalak eltávolítását morfológiai szűrő segítségével hajtom végre.

A morfológia alaktant jelent, és egyaránt alkalmazzák a biológiában és a matematikában is. A matematikában halmazokon elvégezhető műveleteket jelent, ezt használja a képfeldolgozáshoz is. A morfológia célja, hogy a struktúrák és formák tanulmányozása révén valamilyen releváns adatot nyerjünk ki a képből. Ezek a műveletek gyorsan implementálhatók és széles körben alkalmazhatók. Az alapműveletek akkor a leghatékonyabbak, ha bináris képen alkalmazzák őket, ezért én is így tettem.

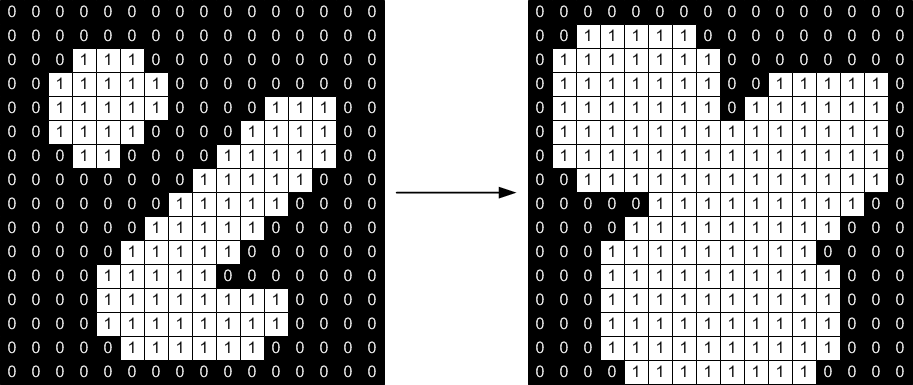
A feldolgozandó kép mellett azonban szükség van egy strukturáló elemre is, amely méretében jóval kisebb a vizsgált képnél, mivel ez fogja végigjárni a kép teljes területét. Ennek a strukturáló elemnek (más néven kernelnek) három különböző alakja lehet. A kernel formája lehet ellipszis, téglalap vagy kereszt.

[kép a 3 kernelről]

Az 1-esek jelölik azt a területet, ahol az adott művelet hatása alkalmazva lesz az adott pixelen, a 0-val jelölt területeket pedig érintetlenül hagyja.

A morfológiai szűrők vagy morfológiai műveletek arra használhatóak fel, hogy kép egyes területeit felnagyítsuk vagy esetleg kicsinyítsük, illetve hogy eltávolítsuk vagy kitöltsük egy terület körüli képpontokat. Az alábbiakban ismertetem ezeket az alapvető műveleteket, amelyek a bemeneti képen lévő objektumokat a kiválasztott strukturáló elemben kódolt jellemzők alapján dolgozzák fel.

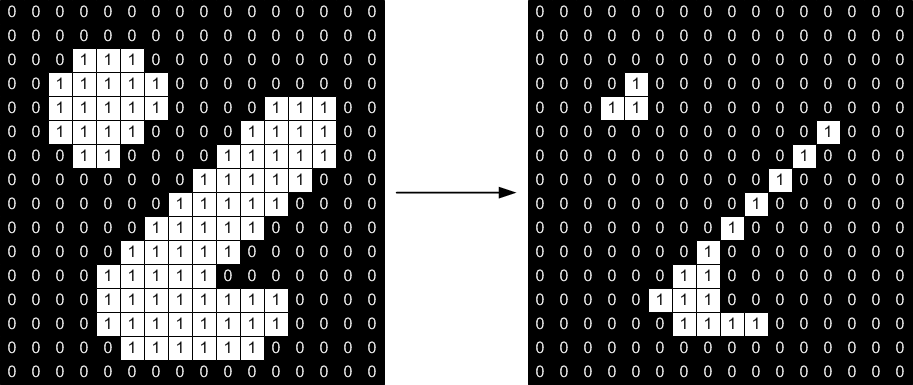
A dilation egyike a két morfológiai alapműveletnek. Ez az operátor azt a hatást éri el, hogy a kép előterében lévő pixelek körüli határokat folyamatosan növeli. Ezzel az előtérben lévő pixelek mérete növekszik, a közöttük lévő lyukak pedig csökkennek. A dilation művelet két adatot vesz be bemenetként. Az első a kép, amelyen a műveletet végre kívánjuk hajtani. A második azon koordinátapontok halmaza, amelyet strukturáló elemként ismerünk. Ez a strukturáló elem határozza meg a dilation pontos hatását a bemeneti képre. A bemeneti kép pixeljeit a bal felső saroktól kezdve megvizsgáljuk a strukturáló elemmel. Minden háttérpixelnél (amit bemeneti pixelnek nevezünk) a strukturáló elemet a bemeneti kép tetejére helyezzük úgy, hogy a strukturáló elem origója egybeessen a bemeneti pixel pozíciójával. Ha a strukturáló elemben legalább egy pixel egybeesik az alatta lévő kép előtérbeli pixelével, akkor a bemeneti pixel az előtér értékére lesz állítva. Ha azonban a képen az összes megfelelő pixel háttérben van, a bemeneti pixel a háttér értékén marad.



8. ábra: Példa 3x3-as téglalap strukturáló elemmel végzett dilation-re

Forrás: <http://www.theobjects.com/dragonfly/dfhelp/4-0/Content/05_Image%20Processing/Morphology%20Filters.htm> (megtekintve: 2021.11.13.)

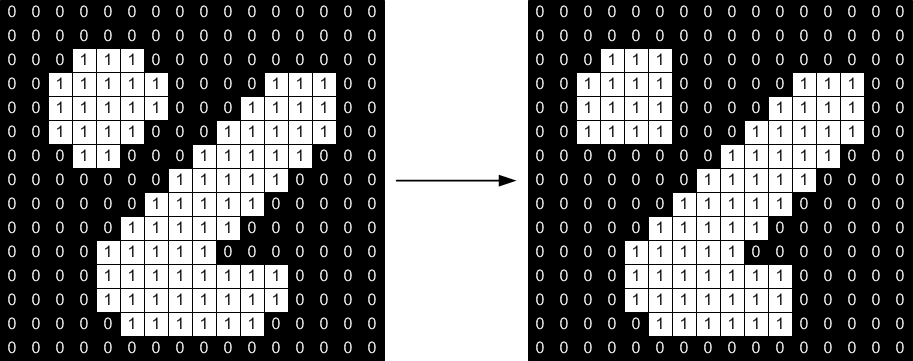
Az erosion a két morfológiai alapművelet közül a másik. Ez az operátor azt a hatást éri el, hogy a kép előterében lévő pixelek körüli határokat folyamatosan csökkenti. Így az előtérben lévő pixelterületek mérete csökken, a lyukak pedig nagyobbakká válnak. A dilation-höz hasonlóan az erosion is két adatot vár bemenetként. Az első a bemeneti kép, a második pedig a strukturáló elem. A művelet végrehajtása is megegyezik az előzőével, azaz a strukturáló elem segítségével a bemeneti kép minden egyes pixeljét megvizsgáljuk, és amennyiben a strukturáló elem minden pixeléhez az alatta lévő kép megfelelő pixelje előtérben lévő pixel, akkor a bemeneti pixel marad úgy, ahogy van. Ha azonban a képen a megfelelő pixelek bármelyike háttér, akkor a bemeneti pixel is háttérértékre lesz állítva.



9. ábra: Példa 3x3-s strukturáló elemmel végzett erosion-re

Forrás: [http://www.theobjects.com/dragonfly/dfhelp/4-0/Content/05\_Image%20Processing/Morphology%20Filters.htm (megtekintve](http://www.theobjects.com/dragonfly/dfhelp/4-0/Content/05_Image%20Processing/Morphology%20Filters.htm%20%20(megtekintve): 2021.11.13)

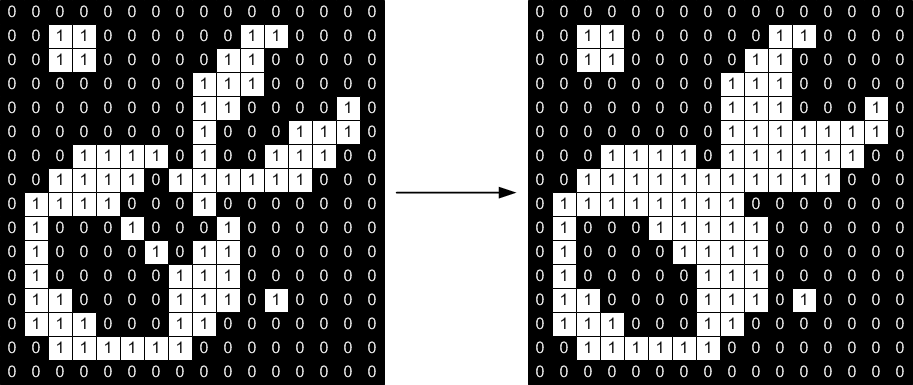
Az opening egy az erosion-ből származtatott operátor. Szintúgy bináris képeken alkalmazzák főként, de akár szürkeárnyalatos képen is működőképes. Az algoritmus hatása az erosion-höz hasonló, ugyanis ennek is az a feladata, hogy az egyes pixelterületek széleit eltűnteti. Azonban ez általánosságban kevésbé ’pusztító’, mint az erosion. Más morfológiai műveletekhez hasonlóan a pontos működést itt is a strukturáló elem határozza meg. Az művelet célja az, hogy megőrizze azokat az előtér-régiókat, amelyek hasonló alakúak ehhez a strukturáló elemhez, vagy amelyek teljesen tartalmazhatják a strukturáló elemet, miközben kiiktatja az előtér pixeleinek összes többi régióját. Nagyon leegyszerűsítve az openinget úgy definiálhatjuk, mint egy erosiont, amelyet egy dilosion követ, mindkét művelethez ugyanazt a strukturáló elemet használva. Míg az erosion elég hatékonyan felhasználható a nemkívánatos előtérbeli pixelek kis csomóinak kiküszöbölésére, pl. 'sózaj', nagy hátránya, hogy az előtérben lévő pixelek minden régióját válogatás nélkül érinti. A nyitás ezt úgy tudja megkerülni, hogy az erosiont és a dilationt is végrehajta a képen. A bemeneti képen elindítja a strukturáló elemet, és csúsztatja körbe az egyes előtér-területeken belül anélkül, hogy megváltoztatná a tájolását. Az összes képpont, amelyet a strukturáló elem lefedhet, és a strukturáló elem teljes egészében az előtérben van, megmarad. Azonban minden olyan előtér pixel, amelyet a strukturáló elem nem tud elérni anélkül, hogy egyes részei kimozdulnának az előtérből, erodálódnak. Az opening elvégzése után az előtérrégiók új határai mind olyanok lesznek, hogy a strukturáló elem beleférjen, így az azonos strukturáló elemmel rendelkező további openingnek nincs hatása.



10. ábra: Példa 3x3-as strukturáló elemmel végzett openingre

Forrás: [http://www.theobjects.com/dragonfly/dfhelp/4-0/Content/05\_Image%20Processing/Morphology%20Filters.htm (megtekintve](http://www.theobjects.com/dragonfly/dfhelp/4-0/Content/05_Image%20Processing/Morphology%20Filters.htm%20%20(megtekintve): 2021.11.13)

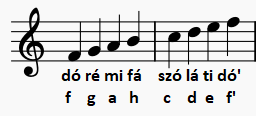
A closing az openinghez hasonlóan az erosion és a dilation alapvető műveleteiből származtatható. Az említett operátorokhoz hasonlóan általában bináris képekre alkalmazzák, bár vannak szürkeárnyalatos verziók is. A closing bizonyos tekintetben hasonlít a dilation-höz, mivel hajlamos megnagyobbítani a kép előtér régióinak határait és csökkenteni a lyukakat az ilyen területeken, de kevésbé roncsolja az eredeti határalakot. A többi művelethez hasonlóan a pontos működést egy strukturáló elem határozza meg. Az operátor hatása az, hogy megőrizze azokat a háttérrégiókat, amelyek hasonló alakúak, mint a strukturáló elem, vagy amelyek teljes mértékben tartalmazzák a strukturáló elemet, miközben kiiktatja a háttér pixelek összes többi régióját. A closing tulajdonképpen egy opening, csak fordított sorrendben. Egyszerűen úgy definiálható, mint egy dilation, amelyet erosion követ, mindkét művelethez ugyanazt a strukturáló elemet használva. A dilation egyik felhasználási módja a képek háttérpixelein lévő lyukainak kitöltése, pl. 'paprika zaj'. Ezzel azonban az egyik probléma az, hogy a dilation a pixelek minden régióját válogatás nélkül megváltoztatja. Azzal, hogy a dilation után a képen erosiont végzünk, ennek a hatásnak egy részét csökkentjük. A kép tetejétől elindul a strukturáló elem az egyes előtér-területeken, és végigcsúsztatja anélkül, hogy megváltoztatná a tájolását. Bármely háttér-határpont esetében, ha a strukturáló elem hozzáérhet a ponthoz anélkül, hogy az elem bármely része egy előtér-területen belül lenne, akkor az a pont háttér marad. Ha ez nem lehetséges, akkor a pixel előtérbe kerül. A closing elvégzése után a háttérrégió olyan lesz, hogy a strukturáló elem a háttér bármely pontját lefedheti anélkül, hogy annak bármely része előtérpontot is takarna, így a további closingnak nincs hatása.



11. ábra: Példa 3x3-as strukturáló elemmel végzett closingra

Forrás: [http://www.theobjects.com/dragonfly/dfhelp/4-0/Content/05\_Image%20Processing/Morphology%20Filters.htm (megtekintve](http://www.theobjects.com/dragonfly/dfhelp/4-0/Content/05_Image%20Processing/Morphology%20Filters.htm%20%20(megtekintve): 2021.11.13)

A top hat operátorok



12. ábra: Hangjegyek által jelölt hangok a kottában (saját szerkesztés)

# Irodalomjegyzék

1. Gary Bradski, Adrian Kaehler: *Learning OpenCV: Computer Vision with the OpenCV Library,* 2008, pp. 153-158.
2. Csatlós István: *Programozási nyelvek kialakulása, fejlődése,* 2014, pp. 1-4.
3. Dagao Duan, Meng Xie, Qian Mo, Zhongming Han, Yueliang Wan: *An improved Hough Transform for line detection,* International Conference on Computer Application and System Modelling, 2010, pp. 354-355.
4. Dumitru Erhan, Christian Szegedy, Alexander Toshev, Dragomir Anguelov, *Scalable Object Detection using Deep Neural Networks,* 2013, pp. 1-3.
5. C. Galambos, J. Matas, J. Kittler: *Progressive Probabilistic Hough Transform for line detection,* Computer Society Conference on Computer Vision and Pattern Recognition, 1999, pp. 554-556.
6. Reagan L. Galvez, Argel A. Bandala, Elmer P. Dadios, Ryan Rhay P. Vicerra, Jose Martin Z. Maningo, *Object Detection Using Convolutional Neural Networks,* TENCON Region 10 Conference, 2018, pp. 2023-2024.
7. Ramesh Jain, Rangachar Kasturi, Brian G. Schunck: *Machine vision,* 1995, pp. 1-31.
8. Prateek Joshi, David Millan Escriva, Vinicius Godoy: *OpenCV By Example,* 2016, pp. 5-10.
9. Kondorosi Károly, Szirmay-Kalos László, László Zoltán: *Objektum orientált szoftverfejlesztés,* 2007, pp. 53-62.
10. Maurício Marengoni, Denise Stiringhini: *High Level Computer Vision Using OpenCV,* SIBGRAPI Conference on Graphics, Patterns, ands Images Tutorials, 2011, p. 13
11. Alexander Mordvinstev, Abid K., *OpenCV-Python Tutorials Documentation,* 2017, pp. 87-106.
12. Maosud Nosrati: *Python: An appropriate language for real world programming,* World Applied Programming, Vol. 1 No. 2, 2011, pp. 110-114.
13. Wang Zhiqiang, Liu Jun, *A Review of Object Detection BAsed on Convolutional Neural Network,* Chinese Control Conference, 2017, pp.11104-11108.
14. https://docs.python.org/3.0/whatsnew/3.0.html (megtekintve: 2021. április 16.)