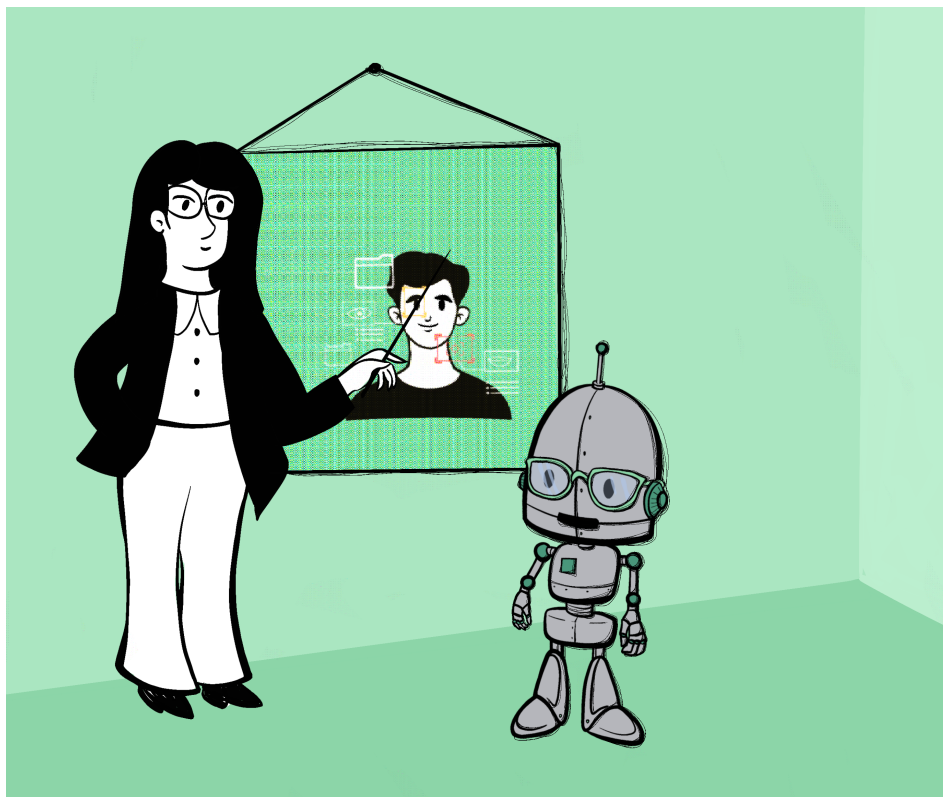


6. Modul

Számítógépes képalkotás

"A **számítógépes képalkotás** összetett terület, ezért még sok kutatásra van szükség ahhoz, hogy az összes ismert - és még ismeretlen - probléma megoldódjon."



A modulról

Ebben a modulban a hangsúly a számítógépes képalkotáson (**CV**) van. Az előző fejezetekben már elsajátított készségeket és ismeretköröket kell kibővíteni, és ösztönözni kell a **CV** témakör kritikai vizsgálatát. A "Felügyelt tanulás" modul ismerete hasznos a későbbi tartalom mélyebb megértéséhez, de nem kötelező annak elvégzése. Az előzetes ismeretek nélküli feldolgozás sem jelent problémát, a legfontosabb tanulási célok e nélkül is érthetőek.

A modul célja, hogy alapvető ismereteket nyújtson arról, hogy miben különbözik az emberek és a gépek "látásmódja", milyen technikák és algoritmusok teszik lehetővé a tárgyak felismerését, de kritikai vizsgálatot és vitaalapot is teremtsen. Ezeket a célokat elsősorban gyakorlati, "kézzel fogható" és "CS-Unplugged" példákon keresztül érjük el, és egy olyan gyakorlati példával zárjuk, amelyben a hallgatók saját gépi tanuláson alapuló arcfelismerő programot fejlesztenek Scratch nyelven. A mesterséges intelligencia mélyebb elméleti ismeretei és matematikai háttérismeretek nem szükségesek.

Célok

A tanulók képesek lesznek:...

- ...megnevezni az emberi és a számítógépes vizuális információszerzés hasonlóságait és különbségeit
- ...elmagyarázni, hogy a számítógép hogyan tárolja és dolgozza fel a vizuális információkat
- ...ismerni a tárgyfelismerés alapvető algoritmusait, és egyszerű példákon keresztül el tudják magyarázni azokat
- ...maguk is alkalmazni az egyszerű gépi tanulási **CV** algoritmusokat
- ...bemutatni a **CV** korlátait és lehetőségeit

Időbeosztás

Idő	Tartalom
10 perc	Bevezetés
15 perc	Elmélet - képfeldolgozás
15 perc	Elmélet - Klasszikus CV eljárások
30 perc	Elmélet + táblázat - Viola Jones

Idő**Tartalom**

40 perc

Gyakorlat - ML arcfelismerés a Scratch segítségével

10 perc

Megbeszélés - lehetőségek és korlátok

Bevezetés – Mi a számítógépes képalkotás (CV)?

A **CV** bevezetője arról szól, hogy a diákok aktívan foglalkoznak a "látás" és a "tárgyfelismerés" alapfogalmaival egy plenáris beszélgetés keretében. A prezentációs diák ennek a bevezető folyamatnak az alapjául és útmutatójaként szolgálnak.

Ember – számítógép összehasonlítás

(2. – 5. dia)

A 2-3. dián a tanulók plenáris beszélgetés keretében megismerkednek azzal a témával, hogy az emberek és a számítógépek hogyan érzékelik a környezetüket. A lényeg itt az, hogy az alapvető folyamat hasonlósága emberek és a számítógépek között láthatóvá váljanak. Különböző módszerekkel (táblák, de akár különböző online eszközökkel is) lehet ötleteket gyűjteni.

Az emberek és a számítógépek is a bemenet – feldolgozás – kimenet elv szerint működnek. A szemünkkel látunk valamit, az agyunkkal feldolgozzuk, és ennek megfelelően cselekszünk. Ugyanakkor ahhoz, hogy e séma szerint tudjunk dolgozni, az embernek (csecsemőkortól kezdve) és a számítógépnek is (pl. felügyelt tanulás) meg kell tanulnia bizonyos készségeket. Ezt egészen addig gyakoroljuk iteratív módon, amíg például ki tudunk mondani egy mondatot, vagy egy almát almaként ismerünk fel. Ezt a kiinduló leckét először az 5. dián mutatjuk be. A tanulási folyamat után már szembe tudunk nézni a kihívással, és beszélgetéseket folytathatunk vagy akár meg tudunk különböztetni egy almát a narancsoktól. A számítógép esetében azonban vizuális anyagot kell biztosítanunk, amelyet a számítógép feldolgozhat. Ez felelne meg a bemeneti lépésnek. Ezt követi a számítógép általi feldolgozás speciális számítógépes képalkotó algoritmusok segítségével. Az utolsó lépés a kimenet, amelynek során például a képen lévő tárgyakat megjelöljük, vagy az arc felismerésével feloldjuk a mobiltelefont.

A 6. dia néhány tipikus **CV** feladatot mutat be.

Besorolás

A képen látható tárgy macskaként való besorolása

Besorolás + lokalizáció

A macska ezen a képen is megtalálható, és piros téglalapokkal jelöljük.

Objektum felismerése

Ebben a példában a képen található különböző objektumok besorolása és lokalizálása is megtörténik.

Tananyag

-  CV - Introduction.pdf

A digitális képek alapjai

Az előző részben a témát alapvetően a megbeszélések és példák segítségével mutattuk be, ezt a tudást a következő részben a digitális képek néhány elméleti alapvetésével kívánjuk elmélyíteni. A hangsúlyt a képek belső feldolgozására és tárolására kell helyezni.

Hogyan tárolódnak a képek a számítógépen?

(8. - 9. dia)

Az információfeldolgozás és -tárolás elvileg binárisan működik. Ez azt jelenti, hogy a számítógép csak a 0 és 1 értéket ismeri. 0 azt jelenti, hogy nem érvényes, 1 pedig érvényest jelent. Most azon kell elgondolkodni, hogyan lehet hatalmas mennyiségű képi információt nullákkal és egyesekkel ábrázolni. Itt a kereten kívül kell gondolkodni. Az egyszerűség kedvéért nézzünk meg egy képet szürkeárnyalatban. Egy szürkeárnyalatos képen a kép különböző területei sötétebbnek vagy világosabbnak tűnnek. Minél sötétebb egy képpont, annál alacsonyabb a reprezentatív számértéke. A világosabb pixelek esetében az érték ennek megfelelően nagyobb. A számtartomány 0 és 255 között van, tehát a 0 nagyon sötét, a 255 pedig nagyon világos. Ha ezt a decimális számot ezután bináris számmá alakítjuk (4 => 1000, 3 => 11), akkor egy számítógépben tárolhatjuk.

A 9. dián az emberi és a számítógépes képalkotás közötti összehasonlítást láthatjuk. Az a mód, ahogyan a képeket a számítógépben tárolják, az egy valóságos adatáradatot eredményez. Ami viszont izgalmas, hogy amint egy kicsit kinagyítjuk ezt a számtömeget, akár a szemünkkel is láthatjuk a számítógépben tárolt technősbéka képének tartalmát pusztán a pixelértékek alapján.

Milyen információkat lehetne már ezekből a pixelértékekből kiolvasni? Ha egy világos négyzet szomszédos egy sötét négyzettel, akkor egy egyszerű kivonással kiderülhet az értékek közötti különbség. Mivel a képeken látható tárgyaknak általában van valamilyen jellegzetes éle, amely színben különbözik a környezetétől, a számítógép egyszerű matematika segítségével felismerheti például, hogy ez egy asztal széle vagy egy focipálya vonala (erről később, az Élek felismerése fejezetben lesz szó).

Hasonlóan működik az RGB-képek esetében is, azzal a különbséggel, hogy sokkal több információt tartalmaz számok formájában. Ez lehetővé teszi a számítógép

számára, hogy kiszámítsa/elmentse a szint, majd megjelenítse azt. Az egyszerűség kedvéért ebben a modulban mindig szürkeárnyalatos képeket vizsgálunk.

Számunkra jelenleg csak az a fontos, hogy ezekből a számhalmazokból speciális algoritmusok és különböző matematikai módszerek segítségével rengeteg információt tudunk kinyerni. Egyszerű bevezetésként közösen megvizsgáljuk, hogyan működik az „áradáskitöltő” algoritmus.

"Áradáskitöltő"

(10. dia)

A 10. dián az „áradáskitöltő” algoritmus működését plenáris megbeszélés keretében kell kidolgozni. A tanulóknak időt adunk arra, hogy átgondolják az algoritmus működését, és leírják a lépéseket. Ezt a fázist követően a lehetséges megoldásokat közösen megvitatják és összehasonlítják.

Hogyan működik az algoritmus? Ahogy a neve is mutatja, a képet "elárasztják" vagy kitöltik/megfestik. A számok itt jó választásnak bizonyulnak a könnyebb szemléltetés érdekében. A képet addig töltjük ki, amíg egy másik szint nem találunk. Attól függően, hogy melyik számot rajzolták, több vagy kevesebb területet festenek. Minden számnak van egy meghatározott szabad, festetlen területe, ezért a számítógép matematikai számítások segítségével azonosítani tudja.

Lépések: (Példa)

1. Vegyünk két szint (háttér és betűszín)
2. Helyezzük az ecsetet egy kezdőpontra
3. Fessünk addig, amíg az ecset alatti pont más színű nem lesz, mint a háttér
4. Amikor minden meg van festve, ellenőrizzük az ecset színének szintjét

Ennek az egyszerű algoritmusnak is számos korlátja van. Mi van például, ha a számokat helytelenül írják, és például a nullának van egy üres helye, és a belső terület is ki van töltve?

Jól ismert példák az algoritmus használatára a legendás Aknakereső játék vagy az MS Paint területkitöltő funkciója.

Tananyag

-  CV - Introduction.pdf

Klasszikus CV algoritmusok

Ezen a ponton a tanulóknak nagyjából el kell tudniuk képzelni, hogy a számítógép hogyan képes egyszerű dolgokat felismerni a digitális képeken. De mi van akkor, ha a tárgyak bonyolultabbá válnak, a színelméletek elmosódottabbak, és az összetettség növekszik?

A tárgyfelismerés gyakran gépi és „mélytanulási” algoritmusok alapján működik, amelyeket az idők során tökéletesítettek. Ezek a nagyon összetett algoritmusok gyakran helyettesíthetők egyszerűbb, klasszikus elődeikkel. A kulcsszó a komplexitás csökkentése az absztrakció révén. Ezt a 14. és 15. dián egy egyszerű példán keresztül szemléltetjük. Még ha minden képi jellemzőt el is távolítunk – az elefánt körvonalának kivételével – az elefánt ezek után is felismerhető marad.

Klasszikus érzékelési algoritmus lehet például a Sobel- vagy a Canny-algoritmus. Ezek az algoritmusok nagyon is a mátrixszámítás matematikai alapjaira épülnek, de az objektumok egyértelmű élei már az algoritmus részlépései után felismerhetők. (16-24. dia)

Gondoljunk vissza a képek tárolásával foglalkozó fejezetre. A 16. dián ismét ezt a belső ábrázolást mutatjuk be példaként. Nagyon jól látható, hogy ez a robot egyes pixelekből áll, amelyek meghatározott számértékekkel rendelkeznek. A könnyebb kezelhetőség miatt a pixelértékek RGB ábrázolását felcseréltük a szürkeárnyaltos ábrázolással, de az elv ugyanaz marad. Ha bizonyos távolságból nézzük a számértékeket, már emberi szemmel is felismerhető egy hozzávetőleges alakzat. De hogyan ismerhetné fel egy számítógép ennek a tárgynak az "élet"? Helyes, matematikai számítások, főként mátrixszámítások révén.

A vízszintes élfelismerés eredménye a következő (17.) dián látható. A bal oldali képet egy Python szkript segítségével készítettük el, kitérőként az érdeklődők számára elérhető. A jobb oldali ábra a vízszintes élszámítás elvégzése után a megfelelő számértékeket mutatja. A 18-19. dián a matematikai technika háttérét részletesebben is elmagyarázzuk. Amennyire bonyolultnak tűnnek ezek az ábrák, annyira egyszerű az érték kiszámításának algoritmusai. Csak alapvető számítási szabályokat alkalmaznak. Például az A_b cellát kivonjuk az A_a cellából, és egy számértéket kapunk. Minél nagyobb ez a számérték, annál nagyobb a színelmélet az egyes pixelek között. Ebből következik, hogy nagy valószínűséggel egy él pixelét ismertük fel. Ha ezt az eljárást az összes pixelértékre

elvégezzük, akkor nem csak egyedi értékeket kapunk, hanem egész vonalakat, jelen esetben éleket, amelyeket felismerhetünk. Most ennek a mátrixnak a további értékeit kell a tanulókkal együtt kiszámolni (a megoldás a megfelelő diák jegyzetek területén található).

A függőleges élek felismerésének elve ugyanígy működik, ezért ezt a továbbiakban nem tárgyaljuk. Ha kiszámítottuk a függőleges és vízszintes éleket, akkor a két eredményt kombinálhatjuk, és így egy olyan mátrixot vagy képet kapunk, amely már tartalmazza az összes élt. Ezt a 23., 24. dián ismét szemléltetjük, és a 25. dián összehasonlítjuk az eredeti képpel.

Dióhéjban az algoritmus a következő három lépésből áll:

- A vízszintes élek kiszámítása
- A függőleges élek kiszámítása
- A két eredmény kombinálása

Tananyag

-  CV - Introduction.pdf

Arcfelismerés a Viola Jones algoritmus segítségével

A 27-30. diát a diákokkal együtt kell megvitatni a közös megbeszélés keretében. Minden szükséges információ megtalálható a diákon.

A 31. dia a Viola Jones algoritmus első technikai alapjait mutatja be. Az algoritmus fontos része a "csúszóablak", amellyel a képet lépésről lépésre balról jobbra és fentről lefelé haladva használjuk, keresve az előforduló (akár többszörös) arcokat és azok vonásait. Az egyes arcrészletek (szem, orr, száj stb.) felismerése úgynevezett "Haar-szerű jellemzőkkel" történik, amelyek a 32. dián láthatóak.

A főbb Haar-szerű jellemzők a következők:




- az éljellemzők
- vonaljellemzők
- négyszög jellemzők

Ezek a jellemzők és a csúszóablak nem rögzített méretűek, hanem szükség szerint dinamikusan módosíthatók. Ez az eljárás a 33. dián ismét látható, majd a 34. dián egy videó is bemutatja, amely összefoglalja és vizualizálja a Viola Jones algoritmus folyamatát.

A Viola Jones algoritmusnak vannak bizonyos korlátai: csak frontális arcokat képes felismerni, ami bizonyos forgatókönyvek és felhasználási helyzetek esetén problémás lehet, és további számítógépes látásmódok alkalmazása szükséges. Azt is fontos hangsúlyozni, hogy bár a Viola Jones algoritmus képes az arcok felismerésére, nem képes megkülönböztetni a különböző arcokat.

A következő (35-36. dia) a Viola Jones algoritmus gyakorlati példája. A szükséges anyagok és a feladat magyarázata megtalálható a feladatlapfájlban és a diákon. Ezen kívül lehetőség van arra, hogy az érdeklődő vagy haladó diákok egy Python szkript segítségével kipróbálják a Viola Jones algoritmust (a feladat és a minta megoldás a 37. dián található).

Tananyag

-  CV - Introduction.pdf
-  <https://youtu.be/hPCTwxFOqf4>
-  CV - Viola Jones.pdf





- Haladó: 📄 CV - Viola Jones Python.zip

Gépi tanulás a CV-ben

A 39. dián egy rövid összefoglalót találunk. A számítógépes képalkotás (**CV**) és a gépi tanulás (ML) az AI két területe. A számítógépes képalkotás a vizuális információk számítógépes feldolgozásával foglalkozik, amelyet gépi tanulási megközelítésekkel tovább lehet fejleszteni (pl. mozgások, tárgysorozatok felismerése valós időben, ...). Mivel a gépi tanulást - különösen a felügyelt tanulást - már egy másik modulban megbeszéltük és tárgyaltuk, a következőkben olyan "gyakorlati" példákra összpontosítunk, amelyekben az ML és a **CV** kombinálódik.

A hallgatóknak egy virtuális okostelefont kell programozniuk egy korábban létrehozott gépi tanulási modell segítségével. A szükséges dokumentumok és magyarázatok a megfelelő munkalapfájlban találhatóak.

Tananyag

-  CV - Introduction.pdf
-  <https://machinelearningforkids.co.uk>
-  CV - Face Lock.pdf
-  CV - Face Lock Solutions.pdf

Lehetőségek és korlátok

A számítógépes képkódolás összetett terület, ezért még sok kutatásra van szükség ahhoz, hogy az összes ismert - és még ismeretlen - problémát meg lehessen oldani. Emellett azt is láttuk, hogy még az olyan egyszerű algoritmusokat is, mint a Viola Jones algoritmus, évtizedek óta gyakorlatilag mindenhol használják. A modul végén a hallgatók esetleges kérdéseire, észrevételeire stb. kell kitérni, valamint kerüljön sor eszmecserére is. A 42. dián található néhány javasolt kérdés, amelyek kiindulópontként szolgálhatnak egy dinamikus beszélgetéshez.

Ha ezen a ponton még van idő, a 43. diát tovább lehet tárgyalni. Itt egy számítógép által besorolt közlekedési lámpa képét változtatják meg egyszerűen a pixelek olyan módon történő megváltoztatásával, hogy a **CV** algoritmus már nem képes helyesen besorolni a közlekedési lámpát. A **CV**-algoritmusok megbízhatóságával kapcsolatos további információk a 43. dián a jegyzetek részben található honlapon találhatóak.

Tananyag

-  CV - Introduction.pdf