

TDK-dolgozat

Lezsák Domonkos

Váladékszemcsék szoftveres kvantifikációja fluoreszcens mikroszkópos felvételek alapján

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

TERMÉSZETTUDOMÁNYI KAR

ANATÓMIAI, SEJT- ÉS FEJLŐDÉSBIOLÓGIAI TANSZÉK



Szerző:

Lezsák Domonkos

programtervező informatikus BSc

3. évfolyam

Témavezető:

Dr. Csizmadia Tamás

egyetemi adjunktus

Lezárva: Budapest, 2022 december

Tartalomjegyzék

1. Bevezetés	4
1.0. Összefoglaló	4
1.1. A dolgozat szerkezete	6
1.2. Biológiai háttér	7
1.2.1. A krinofágia	8
1.3. Technikai háttér	9
1.3.1. Fluoreszcens markerek	9
1.3.2. A GlueFlux rendszer	10
1.3.3. Képalkotás	11
1.3.4. Tipikus kísérlet, összefoglalás	14
1.4. Célkitűzések	14
1.4.1. Milyen a jó statisztika?	16
1.4.2. Felvétel készítés	17
1.4.3. Célkitűzések újraértelmezve	17
1.5. Kapcsolódó munkák	18
1.5.1. Imaris	18
1.5.2. Volocity	19
1.5.3. ImageJ	19
1.6. Elméleti háttér	20
2. Megvalósítás	21
2.1. Szoftveres eszközök	21
2.1.1. Python, mint tudományos nyelv	21
2.1.2. Saját könyvtár — pipeline	22
2.1.3. Saját rendszer — dashboard	24
2.1.4. Elm és V8	24
2.2. Algoritmus	24

2.2.0.	load	24
2.2.1.	clean	27
2.2.2.	diff	30
2.2.3.	sum	30
2.2.4.	find granule centers	31
2.2.5.	analyze coordinates	32
2.2.6.	count	32
3.	Az eredmények értékelése	35
3.1.	Statisztikai validálás	35
3.1.1.	Validálás szakértői kvantifikáció alapján	35
3.1.2.	Validálás biológiai funkciók segítségével	36
3.2.	Sikertelen kísérletek	40
4.	További kutatási lehetőségek	45
5.	Összefoglalás	47
5.1.	Egyéni hozzájárulás	48
A.	Elméleti háttér	50
A.1.	Aritmetikai műveletek képeken	50
A.1.1.	Műveletek kép és skalár között	50
A.1.2.	Műveletek képek között	51
A.2.	Limitálás (thresholding)	53
A.2.1.	Otsu-féle hisztogram alapú threshold	53
A.2.2.	Lokális adaptív limitálás	54
A.3.	Konvolúció	54
A.4.	Gauss-elmosás	54
A.4.1.	<i>Difference of Gaussians (DoG)</i>	57
A.5.	Egyéb folt-detektálási algoritmusok	58
A.5.1.	<i>Laplace of Gaussians (LoG)</i>	58
A.5.2.	<i>Difference of Hessians (DoH)</i>	58
A.5.3.	Mintafelismerés	58
A.6.	Morfológiai módszerek	59
A.6.1.	Erózió, diláció	59

A.6.2. Opening, closing	60
B. Felhasználói dokumentáció	62
B.1. Szoftverkövetelmények, letöltés	62
B.1.1. Futtatás Windows alatt	62
B.1.2. Futtatás Linux alatt (csak fejlesztőknek javasolt)	62
B.2. Fájlformátumok	63
B.3. A program használata	63
B.3.1. Fő program	63
B.3.2. dashboard	68
Irodalomjegyzék	72
Ábrajegyzék	78
Szójegyzék	85

1. fejezet

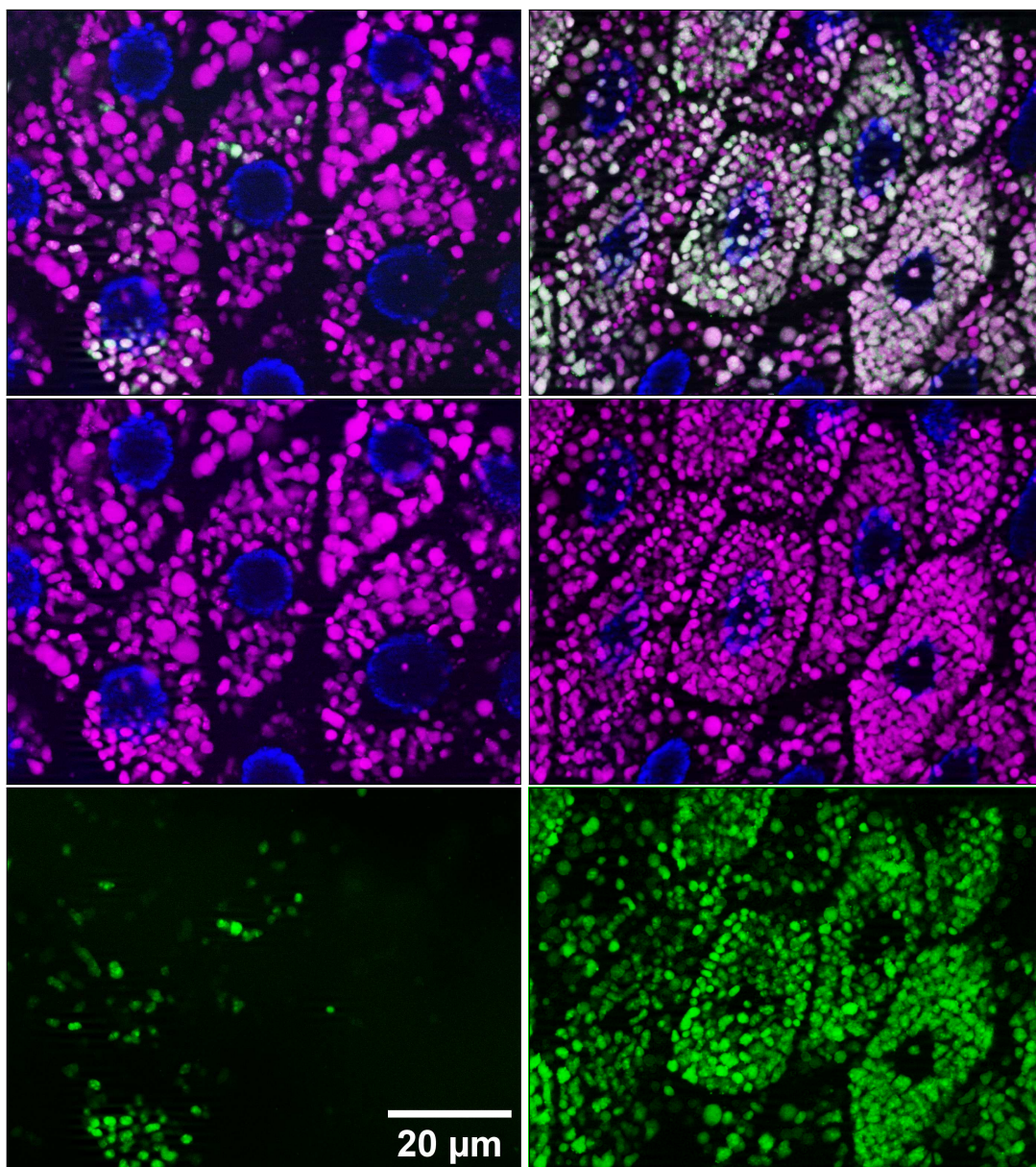
Bevezetés

1.0. Összefoglaló

„A lizoszomális rendszer bármilyen hibája súlyos következményekkel járhat: számos betegség alakulhat ki, amelyek rendszerint súlyos kimenetelűek.” [1] A lizoszomális rendszer egyik feladata a szükségtelenné vált *szekréciós granulumok* (váladékszemcsék) lebontása. Ez egy, a mirigysejtekben természetesen végbemenő folyamat, azonban jelenlegi ismereteink szerint hibája köthető az I-es és a II-es típusú cukorbetegséghez [2], túlműködése pedig a hasnyálmirigy-gyulladásához [3] [4] és a folyamat részletei még mindig nincsenek feltárva.

A folyamat kutatásának egyik részproblémája az 1.1 ábrán látható, ahol mikroszkópos felvételeken meg kell számolni a magenta- és fehér színű foltokként megjelenő granulumokat, ezek arányából következtethetünk bizonyos kezelésekre vagy mutációk biológiai jelentőségére.

Habár lényeges eltérések két kép között ránézésre is megállapíthatóak, a foltok számszerűsítése elengedhetetlen a megalapozott következtetések levonásához. Jelenleg a foltok (granulumok) számlálása kézzel történik, ami kísérletenként több ezer apró döntést jelent. A foltok morfológiájából és elhelyezkedéséből adódóan nem ismerünk olyan módszert, amely lehetővé tenné a folyamat automatizálását, ezért készítettem a kettős célú, bioinformatikai témájú dolgozatomat. Egyrészt szeretnék közvetlenül hozzájárulni több biológiai kutatás hatékonyságának a növeléséhez, másrészt a témakörben egy áttekintést nyújtva szeretnék hasonló fejlesztéseket inspirálni.



1.1. ábra. A kvantifikációs probléma, a különböző színű foltok számolása. Oszloponként ugyanannak a háromszínű felvételnek három különböző projekciója látható: A felső képen az összes szín látható, amit kettébontottam: A második sorban csak a magenta (*szekekrációs granulumokat* jelölő) és a kék (sejtmagokat jelölő) színkomponens, az alsó sorban azonban csak a zöld (*pozitív szekekrációs granulumokat* jelölő) színkomponens szerepel. Az ábráról szemrevételezéssel is leolvasható, hogy a bal oldalhoz képest a jobb oldalon több magenta színben látható folt zöld színben is látszik. A kutatásom fő kérdése: Pontosan hány darab folt (*szekekrációs granulum*) látható, és ezek közül mennyi zöld (kísérlet szempontjából *pozitív* minta) is és mennyi csak magenta színű (kísérlet szempontjából *negatív* minta)?

(A felvételek fluoreszcens mikroszkóppal készültek kontroll (bal oldal) és géncsendesített ($Vps41^{RNAi}$ -kezelt) (jobb oldal) fiatal *Drosophila melanogaster* bábok nyálmirigy-sejtjeiről. A magenta szín DsRed, a zöld szín GFP és a kék szín DAPI jelét mutatja [5])

Az 1.1 ábrán egy szembetűnő eltérés van a bal- és a jobb oldali felvétel között: Jobb oldalon jóval több zöld színt látunk. Ahhoz, hogy el tudjuk dönteni, hogy a foltok számában lévő különbség szignifikáns-e, szükség van a képek számszerűsítésére. Ennek egy kézenfekvő módja, ha megszámloljuk a foltokat, és megszámloljuk, hogy ezek közül hány látható csak magenta színben, és hány zöld (együtt fehér) színben is. Ez a számolás kézzel végezve meglehetősen fárasztó és hosszadalmas munka, de nehezen automatizálható, mivel a legtöbb felvételen a szabálytalan alakú *szekeciós granulumok* jóval kisebbek és jóval több esetben átfedésben vannak egymással. Célszoftver hiányában a kutatók sokszor még mindig ezt az utat kénytelenek választani.

A dolgozat egy olyan *sakértői modellt* állít fel, mellyel elvégezhető a *szekeciós granulumokról* készült többcsatornás mikroszkópos felvételeken a fluoreszcens fehérjékkel megjelölt váladékszemcsék minőségi és mennyiségi elemzése a csatornák közötti összefüggés figyelembevételével.

1.1. A dolgozat szerkezete

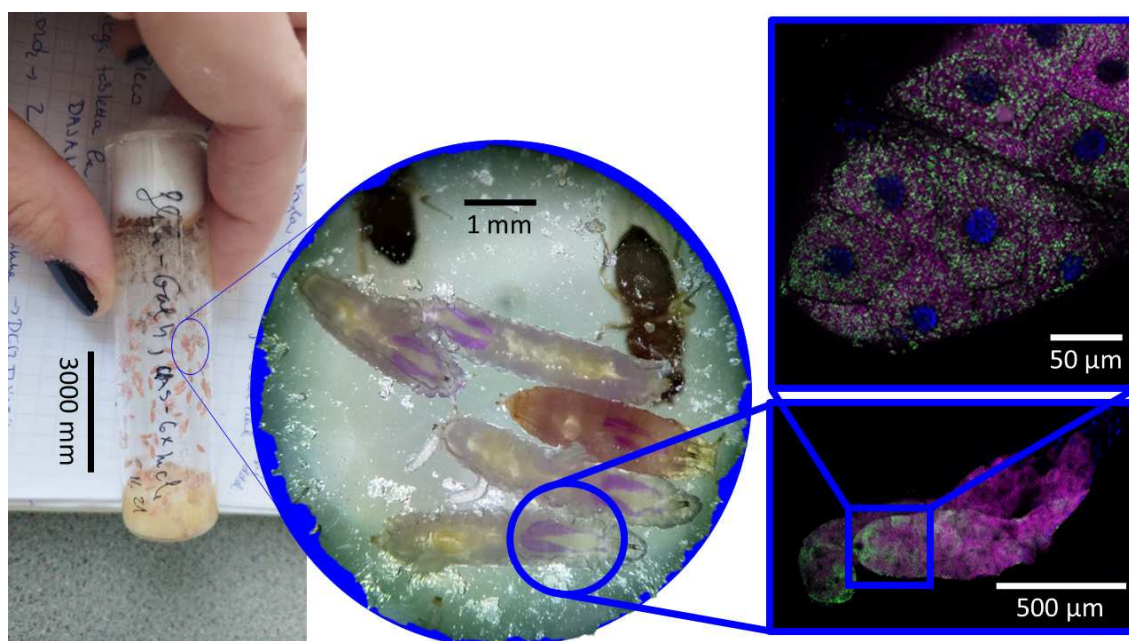
A dokumentumban a dőlt betűvel szedett kifejezések az utolsó oldalakon lévő szójegyzékben magyarázattal ellátva megtalálhatóak, megfelelő PDF-olvasó esetén kattinthatóak, úgy mint az ábraszámok, fejezet/alfejezet/rékszámok és hivatkozások.

A témakör kevésbé kutatott, ezért a munkám jelentős részét a valós probléma (illetve a kapcsolódó problémák) és a lehetőségek felderítése tette ki, ezzel összhangban a bevezetésem (1. fejezet) is hosszú. A bevezetés során először az 1.2 alfejezetben bemutatok részletesebben egy gyakorlatban megoldandó problémát, az 1.3 alfejezetben részletezem a probléma megoldásához használt laboratóriumi eszközöket/eljárásokat, majd az 1.4 alfejezetben összefoglalom a probléma megoldásához kidolgozott *sakértői modellre* vonatkozó elvárásaimat, végül az 1.5 alfejezetben felsorolom a jelenleg létező megoldásokat és a hasonló munkákat. A modellben használt informatikai eszközöket röviden felsorolom a 1.6 alfejezetben, ezeket az A. mellékletben fejtem ki.

A modell részleteit a 2. fejezetben tárgyalom. Kiemelendő, hogy a 2.2 alfejezetben bemutatom a modellem algoritmusát. A dolgozat végén a 3. fejezetben bemutatom az eredményeit, a 4. fejezetben felsorolok pár további fejlesztési lehetőséget, az 5. fejezet pedig egy áttekintő összefoglalást tartalmaz az egész munkáról.

1.2. Biológiai háttér

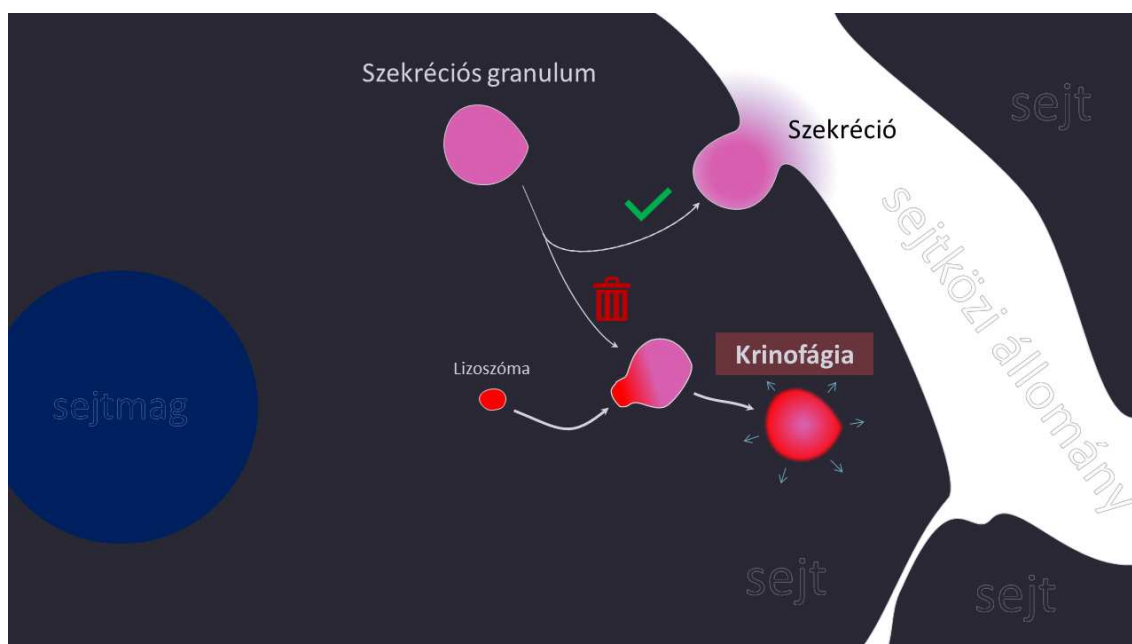
A *Drosophila melanogaster*, vagyis az ecetmuslica előszeretettel használt modellállat genetikai és sejtbiológiai kutatásokban a gyors reprodukciós rátájának és alacsony fenntartási költségeinek köszönhetően. Mindemellett sok olyan folyamat is modellezhető benne, melyek az emberekben is hasonlóan történnek, így a muslicákkal végzett kutatások eredményei később akár gyógyszerfejlesztési kutatások alapjául is szolgálhat. Az egyik ilyen kutatás a *krinofágia* kutatása, ami legkönnyebben a muslica lárváinak nyálmirigyében figyelhető meg. [6]



1.2. ábra. A kísérleti modellállat bemutatása. Balról jobbra sorban először egy muslica fenntartására szolgáló üvegcsövet, majd abban pár kifejlődés előtt álló lárvát, utána lárvák nyálmirigyeit láthatjuk különböző nagyításokban. A képeken lévő állatok transzgenikusak, nyálmirigyeik magenta színűnek látszanak, ami a középső képen szabad szemmel is kivehető. A jobb szélső, nagyobb nagyítású képek hasonlóak, de azokon csak a *szekeciós granulumok* beltartalma került géntechnológiával megjelölésre, így a granulumok elkülöníthetőek, cserébe a magenta szín csak speciális mikroszkópon keresztül látható. Néhány granulum zöld színben is feltűnik egy másik, mesterséges jelölésnek köszönhetően.

(A kék vonalak tájékoztató jellegűek, mindegyik képen másik egyedről készült felvétel.)

A fekete méretvonalal jelölt képeken *+/fkh-Gal4,UAS-6XmCherry* transzgenikus *Drosophila melanogaster* lárvák láthatóak (háttérben sötét színnel két fiatal, kifejlett példány), ennek köszönhetően a nyálmirigyeikben lévő granulumok magenta színben világítanak már látható fényben is. A fehér méretvonalal jelölt képeken hasonló (*GlueFlux* jelzőrendszerrel ellátott) lárvák kiboncolt nyálmirigye látható fluoreszcens mikroszkóp alatt, aminek az esetében a kibocsájtott fény valójában 575 nm-es hullámhosszon a legerősebb, a magenta színt csak a kontrasztarány növelése érdekében végzett szoftveres utószínezésnek köszönhetően kapjuk.)



1.3. ábra. A *krinofágia* folyamata leegyszerűsítve. A szekrécións granulumok a szekréción (normális folyamat) helyett a *krinofágia* útjára is léphetnek [9]

1.2.1. A krinofágia

Az általunk vizsgált *szekrécións granulumok* *Glue*-t tartalmaznak, amely egy ragasztószerű anyag, amit a muslialárvák önmagüknak környezetéhez való rögzítéséhez használnak [7] [8]. A sejten belülről származik, normális esetben kiürül a sejtek közötti állományba (szekretálódik), onnan kerül a nyálba. Amennyiben nem ürül ki a sejtől és a bevezető összefoglalóban említett módon lebomlik, *krinofágiáról* beszélhetünk.

A *krinofágiát* először 1966-ban Smith és Farquhar dokumentálta. Ez egy természetes folyamat, amely során a szekretálásra nem kerülő, abnormális *szekrécións granulumok* (váladékszemcsék) köré kisebb *lizoszómák* (lebontó enzimeket tartalmazó sejtszervecskék) toborozódnak, azzal egyesülnek, majd beltartalmuk elsavasodik és lebomlik (1.3 ábra). [6]

A *krinofágia* működésének részletei (a *szekrécións granulumok* pontosan hogyan, mikor és miért kerülnek lebomlásra) még mindig nincsenek teljesen tisztázva, így ez a témakör a biológia tudományágán belül alapkutatásnak minősül, és nem megfelelő működése számos súlyos problémához vezethet [2] [3] [4] [10], amelyekre vonatkozóan jelenleg is aktív kutatások folynak.

Az 50–100µm átmérőjű sejten belül a vizsgálandó granulumok átlagos átmérője a folyamat során 3 és 8µm között változik (1.2 ábra). A bevezetőben mutatott 1.1 ábrán

magenta színnel ilyen granulumok láthatóak, az ennél kisebb struktúrák (0.1–0.5 μ m átmérőjű lizoszómák) fluoreszcens technológiával közvetlenül nem vizualizálhatóak. Ez a mérettartomány (a látható fény hullámhosszából adódóan) a hagyományos mikroszkópok elméleti felbontásának határát súrolja, ilyen kis méretben videofelvételek készítése óriás kihívás [11] [12], így a kvantifikációs feladat legtöbbször rossz minőségű és álló, de nagy mennyiségű képre irányul. Az egyre növekvő igény miatt jövőbeni terveim közé tartozik a kidolgozott algoritmusomat és a hozzá tartozó implementációt videomikroszkópiára adaptálni.

1.3. Technikai háttér

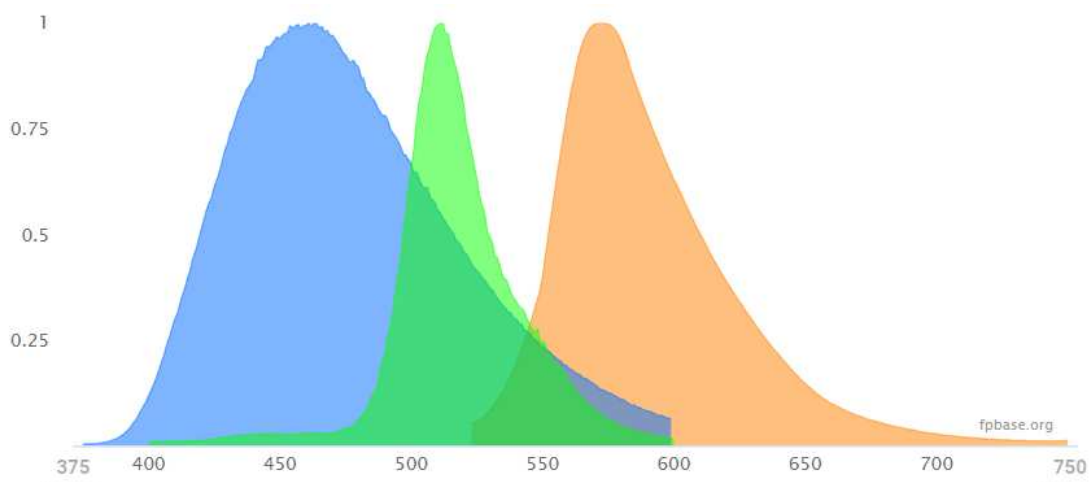
A *krinofágia* a granulumok beltartalmának (esetünkben *Glue*) elsavasodásával jár (a savas közeg az 1.3 ábrán piros színnel jelölve), így — bizonyos feltételek ellenőrzését követően¹ — a *krinofágia* felismeréséhez detektálható a savasodás. Erre megfelelő a GFP (Green Fluorescent Protein) *fluoreszcens marker*, mivel az könnyen felhasználható biológiai kísérletekben és savas közeg hatására elveszíti fluoreszcens képességét (1.3.2 rész). [13]

1.3.1. Fluoreszcens markerek

A *fluoreszcens marker* olyan mesterséges anyag, amely egy bizonyos hullámhosszon történő megvilágítás hatására egy másik, alacsonyabb energiaszintű, de szintén jól meghatározott hullámhossz-tartományban fényt bocsájt ki (fluoreszkál). Az 1.4 ábrán három különböző ilyen anyag spektruma látható. Pontosan ilyen paraméterekkel rendelkező fluoreszcens anyagok a természetben nem jellemzőek. Ha mikroszkóp használata során megfelelő színszűrőt alkalmazunk a fényforrás és a minta között, továbbá a minta és az okulár között, akkor lényegében csak egy olyan *fluoreszcens marker* látható a lencsén keresztül, amivel az állatot mesterségesen elláttuk². Ez lehetővé teszi ugyanabban a mintában található több jelölőanyag szeparált megfigyelését.

¹Egy ilyen feltétel a lizoszómákkal történő fúzió, amelynek detektálásához a LampRing *fluoreszcens markerrendszer* használható. Ez nem képezi közvetlen tárgyát a dolgozatnak, csak a végén fogok róla pár szót ejteni. (Az általam kidolgozott *szakértői modell* csak egy részfeladatát oldja meg a fúzió-detektálásnak.)

²Az állatokat a *fluoreszcens markerekkel* géntechnológiai eljárásokkal lehet ellátni, ennek köszönhetően élő állatok is vizsgálhatóak



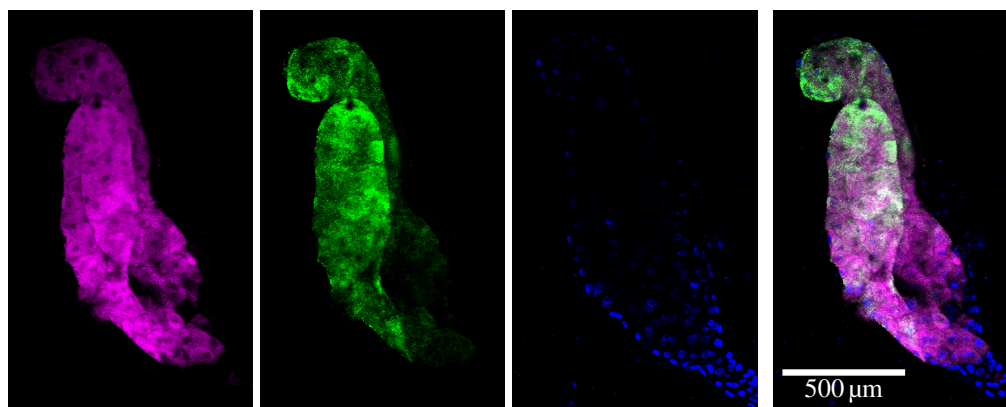
1.4. ábra. Az általam megfigyelt kísérletek során leggyakrabban használt *fluoreszcens markerrendszer*. Kék, zöld és narancssárga színnel rendre a DAPI, GFP és DsRed jelölőanyagok fénykibocsátási spektruma látható egy diagramon ábrázolva. Az én képeimen csak ez a rendszer van használatban, más kísérletek esetén más *fluoreszcens markerrendszer* is előfordulhat. A grafikon az FPbase Spectra Viewer segítségével készült [14]. (A vízszintes tengelyen a fény hullámhossza szerepel 375-től 750 nm-ig, a függőleges tengelyen pedig az egyes anyagok fluoroflorjai által az adott hullámhosszon kibocsátott fény (normalizált) intenzitása látható)

Mivel a minta külső megvilágítása és a fluoreszkálás más hullámhosszon történik, az okulár előtt lévő szűrő csak a *fluoreszcens markerek* által kibocsátott fényt engedi át, így a háttér fekete marad. Így egyszerre több, különböző marker használatával akár színes ábrákat is készíthetünk, ha a különböző megvilágításban készített képeket utólag kombináljuk (1.5 ábra). Ilyen felvételeket fluoreszcens mikroszkópokkal lehet készíteni.

1.3.2. A GlueFlux rendszer

A *GlueFlux jelzőrendszer* egy olyan *fluoreszcens markerrendszer*, amelyben a magentában fluoreszkáló DsRed fehérje (Discosoma Red Fluorescent Protein) és a zöldben fluoreszkáló GFP fehérje (Green Fluorescent Protein) is hozzá van kötve a *szekréciós granulumok*ban fellelhető *Gluehoz*, így az a *Glue*, amely nem savas (semleges vagy lúgos pH-értékű), fehér (zöld+magenta) színben látszik, amely azonban elsavasodott³, csak magentában, így segítségével szépen nyomon követhető a *krinofágia* folyamata. [13] Az 1.6 ábra ennek a rendszernek a működését prezentálja.

³A GFP reverzibilisen inaktiválódik, így ha egy granulum újra visszanyeri semleges pH-ját, újra fehérnek (zöldnek) látszik

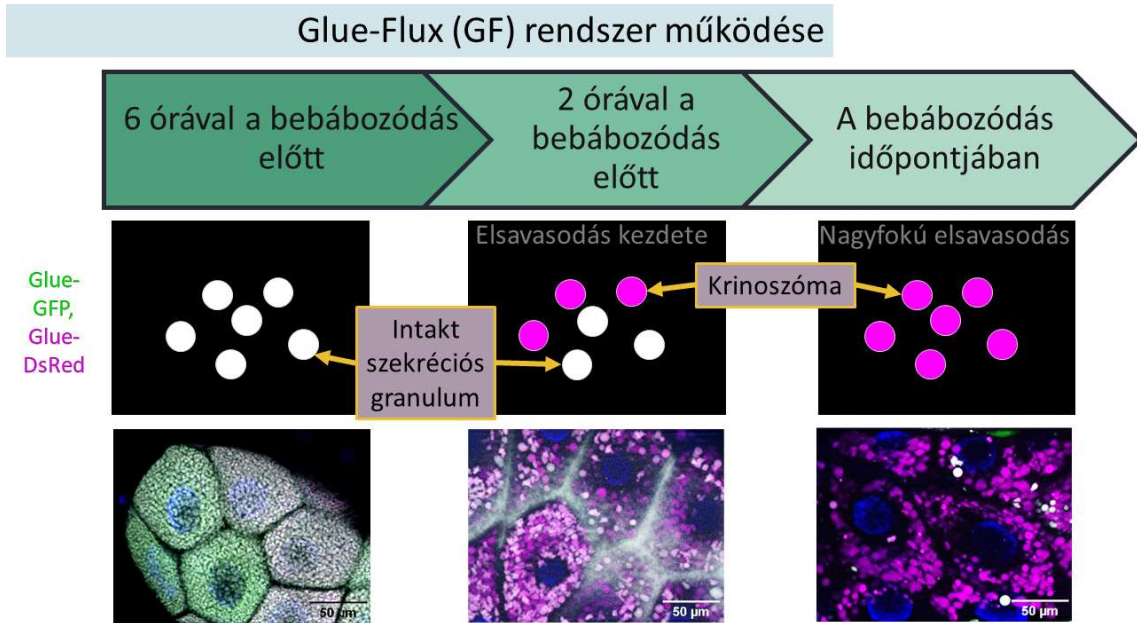


1.5. ábra. *Fluoreszcens* markerek egyesítése. *Drosophila melanogaster* nyálmirigypár, az 1.4 ábra szerinti *fluoreszcens* markerrendszerben: DsRed magenta, GFP zöld és DAPI kék színnel ábrázolva, először színenként, majd egyesítve. A háttér fekete marad, mivel a minta megvilágítása mindig más hullámhosszon történik mint amilyen színben a jelölőanyagok fluoreszkálnak

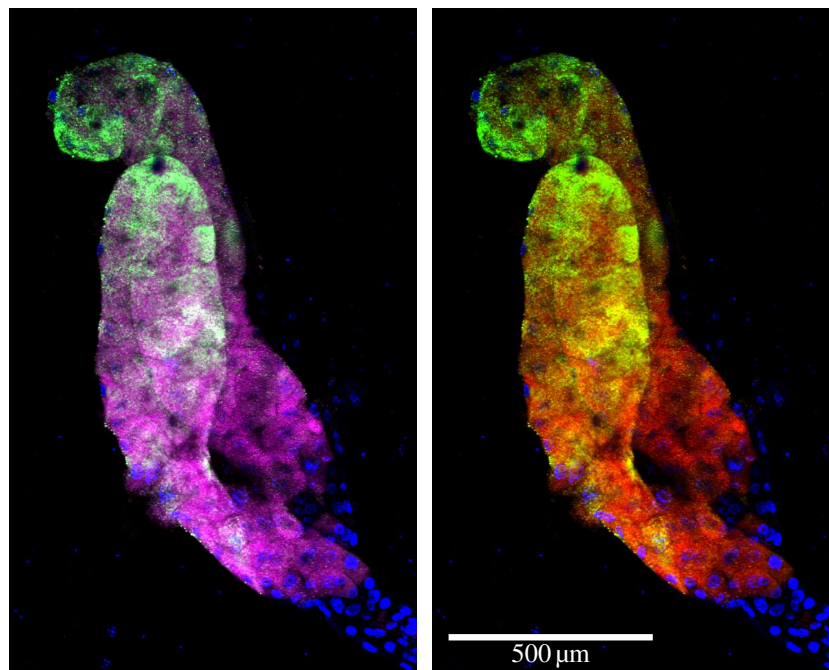
A továbbiakban az általánosíthatóság érdekében a DsRed jelét általában a vörös (rgb(255, 0, 0), eddigiekben magentával volt jelölve), a GFP jelét zöld (rgb(0, 255, 0), ahogy eddig is jelölve volt) színnel fogom megjeleníteni. Ilyen esetben — feltéve, hogy a biológiai rendszer minden tekintetben az elképzelések szerint viselkedik — a semleges belső közegű granulumok sárga (rgb(255, 255, 0)) színben látszanak, míg a savasak vörös színben (rgb(255, 0, 0)). A rendszer mellett a könnyebb tájékozódás érdekében szokás még alkalmazni DAPI *fluoreszcens* markert, amely kék színnel a sejtmagokat jelöli ki. (1.7 ábra)

1.3.3. Képalkotás

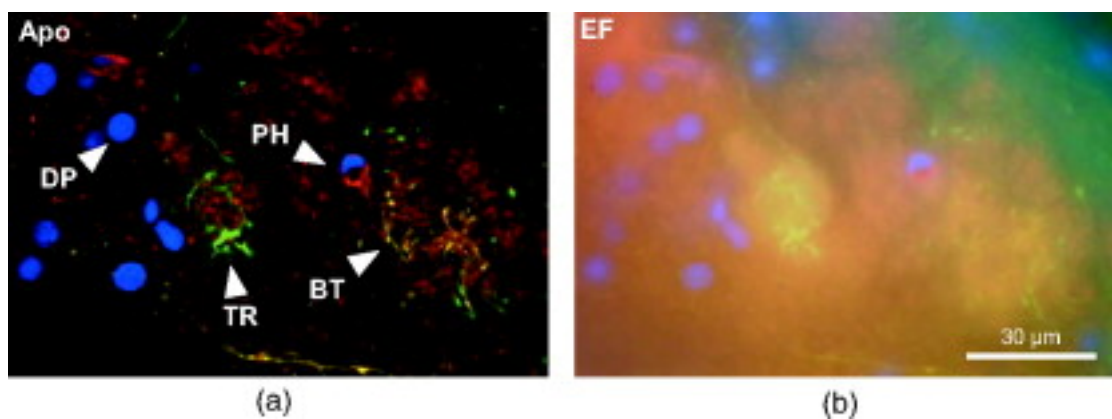
Az 1.5 ábrán a nyálmirigyeknek csak egy síkmetszeti képét láthatjuk. Az ilyen felvételek előállítására a konfokális mikroszkópok képesek, amelyek a fókusz távolság alapján képesek (az objektívre merőlegesen) kizárólag egy optikai szeletről érkező felvételeket mutatni. Az 1.8 ábrán egy összehasonlítás látható optikai szeleteléssel és anélkül készült képekről. [15] Ennek a módszernek az elsődleges előnye az, hogy nincs szükség a minta mechanikus szeletelésére és ezáltal szétroncsolására.



1.6. ábra. A *GlueFlux* jelzőrendszer működése. Az idő előrehaladtával a fehér színű granulumok száma csökken, és egyre több magentává válik, ezt a jelenséget a granulumok beltartalmának elsavasodása okozza. A középső sorban sematikus ábrák, az alsó sorban pedig valódi mikroszkópos felvételek láthatóak. A kép forrása Dósa Anna 2022-es biológia TDK előadása [9], némi átalakítással



1.7. ábra. Magenta-zöld-kék és vörös-zöld-kék színezés összehasonlítása. Bal és jobb oldalon ugyanaz az információ látható, azonban jobb oldalt könnyebb összehasonlítani az első két csatornát.



1.8. ábra. Apotome és szeletelés nélküli (Widefield) képkészítési technológiák összehasonlítása. A bal oldalon a nyíllal jelölt objektumok jobban kivehetők annak köszönhetően, hogy nincsenek egyéb rétegből származó zavaró jelek.
(„Egy 70µm vastag szelet *X. laevis* ebihalról. Kék: sejtmag (DAPI [DP]), zöld: idegek (Alexa Fluor 488 lekövetés (tracing [TR])), sárga: dendritek (Alexa Fluor 546 biocytin visszakövetés (backtracing [BT])), vörös: aktin-szálak (Alexa Fluor 647 phalloidin immunfestés [PH]). (a) Apotome képkészítéssel és (b) szokásos epifluoreszcens kép.”
Az ábra forrása Weigel, Schild és Zeug munkája [15])

1.3.4. Tipikus kísérlet, összefoglalás

A *krinofágia* kutatására irányuló kísérletek egy tipikus folyamata a következő:

Biológiai szempontból vizsgálandó fehérjék kiválasztása (Csizmadia és tsai. [13]) és *fluoreszcens markerrendszer* megtervezése (1.3.1 *Fluoreszcens markerek*)

Mely fehérjéhez milyen színben világító *fluoreszcens markerek* fognak kapcsolódni, és ezek jelenléte milyen biológiai funkciókhoz köthető?

Kísérleti kezelés elvégzése⁴, a minták előkészítése⁵ (nem része a dolgozatnak)

Megfigyelés (1.3.3 *Képalkotás*)

Optikai szeletelésre alkalmas fluoreszcens mikroszkóp használatával

Dokumentálás (1.4 *Célkitűzések*)

Felvételek készítése és elmentése, majd kvantifikációja az egyszerűbb feldolgozhatóság érdekében

Statisztikai ellenőrzés (B.3.2 *dashboard*)

A kinyert adatokra statisztikai mutatók számítása

Publikálás (nem része a dolgozatnak)

Természetesen a valóságban a folyamat ennél jóval bonyolultabb, csak az automatizált kvantifikáció szempontjából lényeges lépések vannak kiemelve.

Habár a dolgozatom során Csizmadia cikkének [13] részleteit veszem alapul, a kidolgozott eljárás jó alapot szolgál más jellegű kísérletekhez is, mint az egy későbbi kutatás [16] során is kiderült.

1.4. Célkitűzések

A fent felvázolt kísérlettípus során felmerül egy gyakorlatias probléma: A *szekrétációs granulumok* elsavasodásának megvizsgálásához meg kell számolni, hogy az összes *szekrétációs granulum* (DsRed foltok) közül mennyi marad semleges pH-jú (GFP foltok, ezeket tekintjük *pozitív* eseteknek), és mennyi savasodik el (ezekben a *negatív* esetekben

⁴A gyakorlatban többnyire a kísérlet tárgyat képező kezelés is géntechnológiát takar, géncsendesítést vagy túltermeltetést

⁵A DAPI jelölőanyag közvetlenül a felvételi készítés előtt, sejtmagfestéssel kerül a mintába. A fluoreszcens mikroszkópok a nagy nagyítás miatt érzékenyek, így a mintákat érdemes fixálni illetve alacsony törésmutatójú glicerinnel oldatba helyezni

kialszik a GFP jel). A konkrét problémára jelenleg létező megoldás nincs, és a hasonló problémák esetén is a pontosságra hivatkozva többnyire a kézi számlálásra hagyatkoznak a sejtbiológia területén folyó kutatásokban. Ezekre egyesével kell rákattintani képanalizáló szoftverben, ami rengeteg időt vesz igénybe⁶. Én meg vagyok győződve arról, hogy kellő pontosság elérhető automatizált módszerekkel is, így a dolgozatom jelentősen hozzájárulhatna több biológiai kutatás hatékonyságának a növeléséhez.

Automatizált információkinyerésre több különböző megközelítés létezik. Modell-vezérelt módszereknek nevezzük azokat az eljárásokat, amik esetében matematikai jellegű összefüggésekre támaszkodva készítünk egy adathalmazból predikciót. Adat-vezérelt módszereknek nevezzük azokat az eljárásokat, amik esetében az adathalmaz matematikai jellegű összefüggéseire tekintet nélkül alkotunk predikciós eljárást. Ennek tipikus esete a mélytanuló algoritmusok használata. A két módszer között félúton helyezkednek el azok az eljárások, amelyek egy modellvezérelt módszerrel jellemzőket emelnek ki az adatokból, amelyeket felhasználnak az adatvezérelt módszerek adataiként, így gyorsítva és javítva az adott esetben alkalmazott mélytanuló (vagy egyéb, gépi tanulós) algoritmus működését [19].

Szerettem volna a korábbi kutatásokra támaszkodva (A. *Elméleti háttér*) egy olyan *szakértői modellt* felállítani, amellyel a fenti felsorolásban (1.3.4 *Tipikus kísérlet, összefoglalás*) a dokumentálást (időigényes számolgatást) automatizálni lehet, amely alkalmas granulumokról készült fluoreszcens képek beavatkozás nélküli⁷ kvantifikációjára.

Annak érdekében, hogy minél nagyobb legyen a munkám hozzáadott értéke, a feladatot némiképp általánosítottam ahelyett, hogy csak egy konkrét megoldandó problémára koncentráltam volna. Így a modellem némi (feladatfüggő) átalakítással alkalmazható jellemzőkiemeléshez gépi tanulási algoritmusok bemeneteként.

⁶A Google első találata az „imagej count objects” kifejezésre [17] *Labno Christine dokumentumára* mutat, ami manuális számlálást és egyszerű limitálás alapú számlálást ír le. Az utóbbi pontatlan (amit később kifejték). Az ImageJ legfrissebb dokumentációjában szereplő leírás is csak ezt a két eljárást említi [18]

⁷A célkitűzéseim közül a beavatkozás nélküli működés csak részben valósult meg. Az eredmény csak sztochasztikusan közelít a valódi eredményhez, azonban az implementációhoz készítettem egy olyan felhasználói felületet, amellyel korábbi időráfordítások töredéke alatt pontosítható a becsült eredmény

1.4.1. Milyen a jó statisztika?

Az elégséges statisztika „minden információt tartalmaz az ismeretlen paraméterre vonatkozóan”, elég annak az értékét ismerni és „nem függ az ismeretlen paramétértől” [20].

A fluoreszcens mikroszkópos felvételekről egy olyan statisztikát szeretnék készíteni, amely alapján meghatározható a *pozitív*nak tekinthető *szelekciós granulumok* aránya az összes *szelekciós granulum*hoz viszonyítva.

Ezt egy kétlépcsős folyamattal oldom meg: Először felismerem az objektumokat (*szelekciós granulumok*), majd a felismert objektumokat — alapvetően — két osztályba sorolom be (*pozitív* vagy *negatív*).

Sajnos a gyakorlatban többször szakértők között is ellentmondás van annak tekintetében, hogy mi az, amit egy felvétel alapján különálló *szelekciós granulum*nak tekinthetünk, illetve ezek közül melyeket tekinthetjük *pozitív* esetnek és melyeket *negatív*nak. Az előbbi problémát talán úgy érdemes áthidalni, ha a felismerések *unióját* vesszük (tehát elég, ha egy személy határoz meg egy foltot granulumként, nem szükséges a szakértők között teljes egyetértés), az utóbbit viszont lehet modellezni egy skalár paraméterrel (elfogadási limit).

Milyen a jó hiba?

A feladat gyakorlati jellege miatt még a fenti, precíz megközelítés ellenére is vannak olyan esetek, amikor nem beszélhetünk helyes válaszról. Van olyan eset, amikor nem tudjuk osztályozni az adott *szelekciós granulumot* külső tényezők miatt, például éppen két granulum készül fuzionálni, vagy túl nagy a háttérzaj.

A felismerhetetlen esetek tekintetében a modellem osztályozásaival szemben a következő elvárásokat támasztom:

- Mérhető

A bizonytalan esetek mennyiségét is meg kell határozni vagy legalább becsülni, hogy kézi felülvizsgálással javítani lehessen azokat.

- Reprezentatív

A bizonytalanak nyilvánított esetek között hasonló arányban forduljanak elő a *pozitív* és a *negatív* esetek.

- Jól definiált

Amennyiben manuálisan megvizsgálunk egy képet (leülünk, és megszámloljuk a granulumokat), az egyes esetekről el lehessen dönten, hogy várható-e a modell tévedése azok tekintetében.

1.4.2. Felvétel készítés

Én a kutatás során a könnyen elérhető Zeiss AxioImager Z1 és M2 fluoreszcens szemikonfokális mikroszkópok⁸ felvételeire optimalizáltam a *szakértői modellemet*. Ezek az Apotome technológiát használják, amely csak részleges optikai szelektálást végez a gyorsaságért és a biológiai értelemben vett kíméletességért cserébe, szemben a CLSM (Confocal Laser Scanning Microscopy) technológiával. Ez a fluoreszcens technológiát használó vizsgálatoknak a minimális feltétele, ezért feltételezhető, hogy a hasonló kísérleteket végző laborokban ma már mindenhol megtalálható hasonló kapacitásokkal rendelkező műszer.

Elérhető több, jobb felvétel készítésére alkalmas technológia a modernebb RCM (Re-scan Confocal Microscope) [21], azonban egyrészt a kutatásommal szélesebb rétegeket szeretnék minél előbb támogatni, másrészt a nagyítás növelésével csökken a felbontás, és a kis méretű objektumok esetén hasonló problémával állunk szemben, mint rosszabb mikroszkóp esetén, így más jellegű kísérletnél szükség lehet az itt kidolgozott eljárások alkalmazására.

1.4.3. Célkitűzések újraértelmezve

Ahogy Csetverikov már 1989-ben is megfogalmazta, a „számítógépes látásnak prediktív erejűnek kell lennie”. [22] Ennek a figyelembevételével írtam a dolgozatomat. A kutatás során sikerült egy olyan, teljes körű *szakértői modellt* kidolgozni, amely statisztikailag igazolhatóan működik, megfelel a fenti (irányadó) feltételeknek, amellet, hogy több kísérlet sajátosságait is figyelembe veszi.

⁸Az ELTE Biológiai intézet Anatómiai Sejt- és Fejlődésbiológiai Tanszék laborjaiban (ahol a kutatásomat végeztem) ez a két mikroszkóp megtalálható, továbbá több, hasonló témakörben született tudományos cikkben is használtak Zeiss AxioImager mikroszkópokat

1.5. Kapcsolódó munkák

Az általam felvázolt feladat megoldható kolokalizáció vizsgálattal [23], azonban a pontosabb, részletesebb és átláthatóbban indokolható eredmény elérése érdekében szeretném diszkretizálni az egyedi objektumokat, így a munkámra alapozva elvégezhető más jellegű kísérletek (például membrányúrré asszociálódó GFP jel keresése, 3.2 alfejezet) kvantifikálása és más jellegű adat kinyerése (például trajektória analízis) is.

Az én kutatásomhoz talán Bradbury 1979-es cikke [24] áll legközelebb. Ebben saját tapasztalatai alapján általános jelleggel bemutatja, hogy a mikroszkópos képek feldolgozásával kapcsolatos kísérletek esetén mire érdemes odafigyelni, mik a buktatók, és mik a javasolt eljárások. A „Problems of grey level discrimination” és „Problems due to image complexity” fejezetében sok praktikát bemutat a képfelismeréses feladatokat illetően, több megoldást, amit ő bemutat, én is kipróbáltam. A Bradbury féle cikk [24] írásának idején még jóval fejletlenebbek voltak a személyi számítógépek, mint ma (lásd „Data handling” című fejezetét), továbbá új eredmények is születtek a képfeldolgozás területén [25] [26], ezért úgy gondolom, lassan időszerű újragondolni az elképzeléseit, tekintettel a mikroszkópos képek feldolgozásának egyre növekvő igényére.

1.5.1. Imaris

Az interneten könnyedén lehet rengeteg olyan cikket, ingyenes programot, vagy tutorial videót találni, ahol mikroszkópos képek kvantifikálásával foglalkoznak, azonban ezek nagy részében nem granulumok, hanem sejtek kerülnek megszámlálásra.

A dolgozatom írása során csak egy olyan programot találtam, amely talán alkalmas lehet a fent vázolt probléma megoldására, az Imarist (Oxford Instruments) [27]. Ez egy kereskedelmi forgalomban lévő, nagyon esztétikus, professzionális és felhasználóbarát szoftver, azonban *szekréciós granulumok* szintjén csak korlátozott megoldásokat kínál. A funkciói közül az 1.2. alfejezetben taglalt problémának a megoldására talán az „Object-Based Colocalization” funkciója áll a legközelebb. Sajnos ennek a funkciónak a pontos működése nem pontosan dokumentált, azonban ha az elképzeléseim helyesek, csak diszkretizáció után képes objektumok közötti összefüggések vizsgálatára, ezzel kizárva bizonyos típusú kvantifikációknak (váratlan területeken lévő jel figyelmen kívül hagyása, granulumon belül csak szórványosan megfigyelhető jel) a lehetőségét.

Rész-probléma	Szoftver Gyártó
Előfeldolgozás	Volocity 6.3 Quorum Technologies, Puslinch, ON, Canada
Objektum felismerés	Volocity 6.3 Quorum Technologies, Puslinch, ON, Canada
Előfeldolgozás	Adobe Creative Cloud Photoshop Adobe, San Jose, CA, USA
Adat-összegzés	Microsoft Office 365 Excel Microsoft, Redmond, WA, USA

1.1. táblázat. Ma és Brill által használt [28] szoftveres megoldások rendszerezve

1.5.2. Volocity

A konkrét problémát talán Cheng-I Ma és Julie Brill meglehetősen friss cikke [28] közelíti meg. Ők a Volocity szoftvert [29] használták mind az előfeldolgozásra, mind az objektum felismerésre (lásd 1.1 táblázat).

A Volocity szoftver [29] nagyon jó beépített *dekonvolúciós* algoritmussal rendelkezik, azonban a *dekonvolúció* nem képi a tárgyát ennek a dolgozatnak.

A *dekonvolúción* túl a Volocity [29] funkciói közül a fájlkezelést leszámítva csak a *Find Objects* funkciót használja ki a Ma és Brill által leírt protokoll [28]. Ez egy limitálást (idegen szóval thresholding, lásd A.2 alfejezet) és utána az objektumok méret alapú heurisztikus szétválasztását takarja. A probléma ilyen jellegű megközelítésére több jól működő megoldás létezik, például Nasr-Isfahani, Mirsafian és Masoudi-Nejad megoldása [30]. Azonban az egyszerű limitálás ilyen jellegű alkalmazását én nem javaslom, mivel rengeteg információvesztéssel jár.

1.5.3. ImageJ

Az ImageJ [31] bioinformatikai területen nagyon gyakran alkalmazott képelemző szoftver, nem véletlenül. Az ImageJ [31] egy ingyenes, szabad és moduláris képelemző szoftver, amely rengeteg plugint tartalmaz. Sajnos ezeknek egyike sem alkalmas önmagában az 1.2 alfejezetben taglalt probléma megoldására, de azokat egymás után alkalmazva⁹ lefuttathatjuk a 2.2 alfejezetben taglalt algoritmust, amivel elérjük a dolgozat által kívánt hatást.

⁹Az ImageJ nagyon jó segédfunkciókkal rendelkezik makróíráshoz, így a grafikus felületen egyesével végrehajtott lépéseket könnyen lehet ismételni, akár programkód olvasása és értelmezése nélkül is

1.6. Elméleti háttér

A dolgozatom elkészítése során a lehető legegyszerűbb algoritmusokra igyekeztem támaszkodni. A morfológiai módszereket részesítettem előnyben, mivel azokra építkezve az algoritmus helyessége intuitívan indokolható [26].

Alább felsorolom azokat a fontosabb eljárásokat, algoritmusokat, amelyeket felhasználok a modellemben; ezeket az A. függelékben részletezem a biológus vagy a témában nem jártas olvasók számára.

- Aritmetikai műveletek kép és skalár között
- Pixelenkénti aritmetikai műveletek két kép között
- Egyszerű limitálás (thresholding) rögzített értékkel
- Gauss-elmosás
- Morfológiai műveletek
 - Erózió, diláció
 - Opening, closing

2. fejezet

Megvalósítás

A fenti módszereket alkalmazva kidolgoztam egy olyan algoritmust, amely képes megoldani a bevezetőben ismertetett feladatot. Az algoritmus egy implementációja elérhető a <https://github.com/lezsakdomi/gfp-ring-detector/tree/GF> linken.

2.1. Szoftveres eszközök

A modellem *magas szintű*, így a megvalósításhoz több külső és belső szoftveres eszközt igénybe vettem. Az implementációhoz készítettem rövidebb felhasználói útmutatót (B. függelék), alább csak a programtervezési szempontból lényegesebb választásaimat részletezem.

2.1.1. Python, mint tudományos nyelv

Az algoritmus implementálását illetően a Python nyelvre esett a választásom, mivel az „nagyszerű” nyelv tudományos problémák prototipizálásához [32]. Alább kiemelek pár szempontot, ami meglátásom szerint ideális választássá teszi azt.

- Gyors prototipizálás
- Rengeteg elérhető adatfeldolgozó könyvtár
- Opcionális kiegészítés gépi kóddal
 - Pár tervezési elvvel (műveletek vektorizálása, számuk csökkentése, adatmásolás elkerülés) és a numpy nevű könyvtár használatával jelentősen

növelhető a teljesítmény, amellet hogy a kód átlátható és könnyen karbantartható marad [33]

- Ha teljes kontrollt akarunk a program futása fölött, a Pythonhoz hasonló szintaxist megtartva kikerülhető az interpretálás [34]
- Akár a programkód lényegi változtatása nélkül is elérhető közel natív teljesítmény JIT (Just In Time compiler) használatával [35] [36]
- Igény esetén saját modulok is készíthetők tetszőleges, alacsonyabb szintű programozási nyelven

2.1.2. Saját könyvtár — pipeline

Az algoritmusom *pipeline*-szerű, tehát jól elkülöníthető, lépcsőzetesen egymásra épülő blokkokból áll.

A kísérletezés során sokszor szükség volt bizonyos kódrészletek kicserélésére, változtatására, áthelyezésére. Ezeknek a változtatásoknak a megkönnyítésére készítettem egy univerzálisan használható, önálló könyvtárat (*pipeline_lib.py*), amellyel könnyen építhetők ilyen *pipeline*-szerű programok. A 2.1 forráskód egy példát mutat a könyvtár használatára, míg a 2.2 forráskód ugyanazt a problémát a könyvtár nélkül oldja meg, kevesebb kiterjesztési lehetőséggel (pl. újra futtatáshoz újra kell indítani az alkalmazást, nincs lehetőség a részképek kinyerésére).

Megjegyzés. A program belső felépítése sok tekintetben hasonlít a CellProfiler [37] felhasználói felületére, ami szétválasztja a képek felderítését, feldolgozását és az adatok feldolgozását.

```
1 class MyPipeline extends Pipeline:
2     @Step.of('DsRed')
3     def add_GFP_to_DsRed(GFP, DsRed):
4         return DsRed + GFP
5
6     @Step.of(['GFP', 'DsRed'])
7     def double(GFP, DsRed):
8         return GFP * 2, DsRed * 2
9
10    @Step.of('summary'):
11    def summarize(GFP, DsRed):
12        return f"Maximum of GFP is {np.amax(GFP)}, " +
13               f"but DsRed peaks at {np.amax(DsRed)}."
```

2.1. forráskód. *Pipeline_lib.py* példa használat

```
1 def add_GFP_to_DsRed(GFP, DsRed):
2     print("Adding GFP to DsRed")
3     return DsRed + GFP
4
5 def double(GFP, DsRed):
6     print("Doubling GFP and DsRed")
7     return GFP * 2, DsRed * 2
8
9 def summarize(GFP, DsRed):
10    print("Summarizing GFP and DsRed")
11    return f"Maximum of GFP is {np.amax(GFP)}, " +
12           f"but DsRed peaks at {np.amax(DsRed)}."
13
14 def run():
15     global _DsRed, _GFP, _summary
16     _DsRed = add_GFP_to_DsRed
17     _GFP, img_DsRed = double(GFP, DsRed)
18     _summary = summarize(_GFP, _DsRed)
```

2.2. forráskód. Példa kód *pipeline_lib.py* használata nélkül

2.1.3. Saját rendszer — dashboard

A prototipizálás későbbi szakaszában az elképzeléseim validálásához komplex adatvizualizációs eszközökre volt szükség. Így elkészítettem egy elkülönített programkomponenst (*dashboard*), amely képes önállóan is futni, azonban sok szempontból kényelmesebb a fő programmal együtt használni. Ez képes soronként kettősponttal jelzett felsorolásokat beolvasni TXT fájlokból, az adatokat CSV formátumú táblázatban elmenteni, különböző szempontok szerint személyre szabható grafikonokon megjeleníteni, intuitív módon módosítani és egy-egy adatpontot a fő programban megnyitni és újraanalizálni. A B. függelék B.4 ábráján néhány képernyőmentés látható a komponens egy tipikus diagramjáról.

2.1.4. Elm és V8

Az implementáció (fő program) felhasználói felületét a hibák minimalizálása érdekében egy funkcionális reaktív programozási nyelven írtam [38]. Az Elm 19.0.1-es verzióját használtam, amely a V8 JavaScript motorral karöltve gyorsan működő weboldalak fejlesztését teszi lehetővé [39].

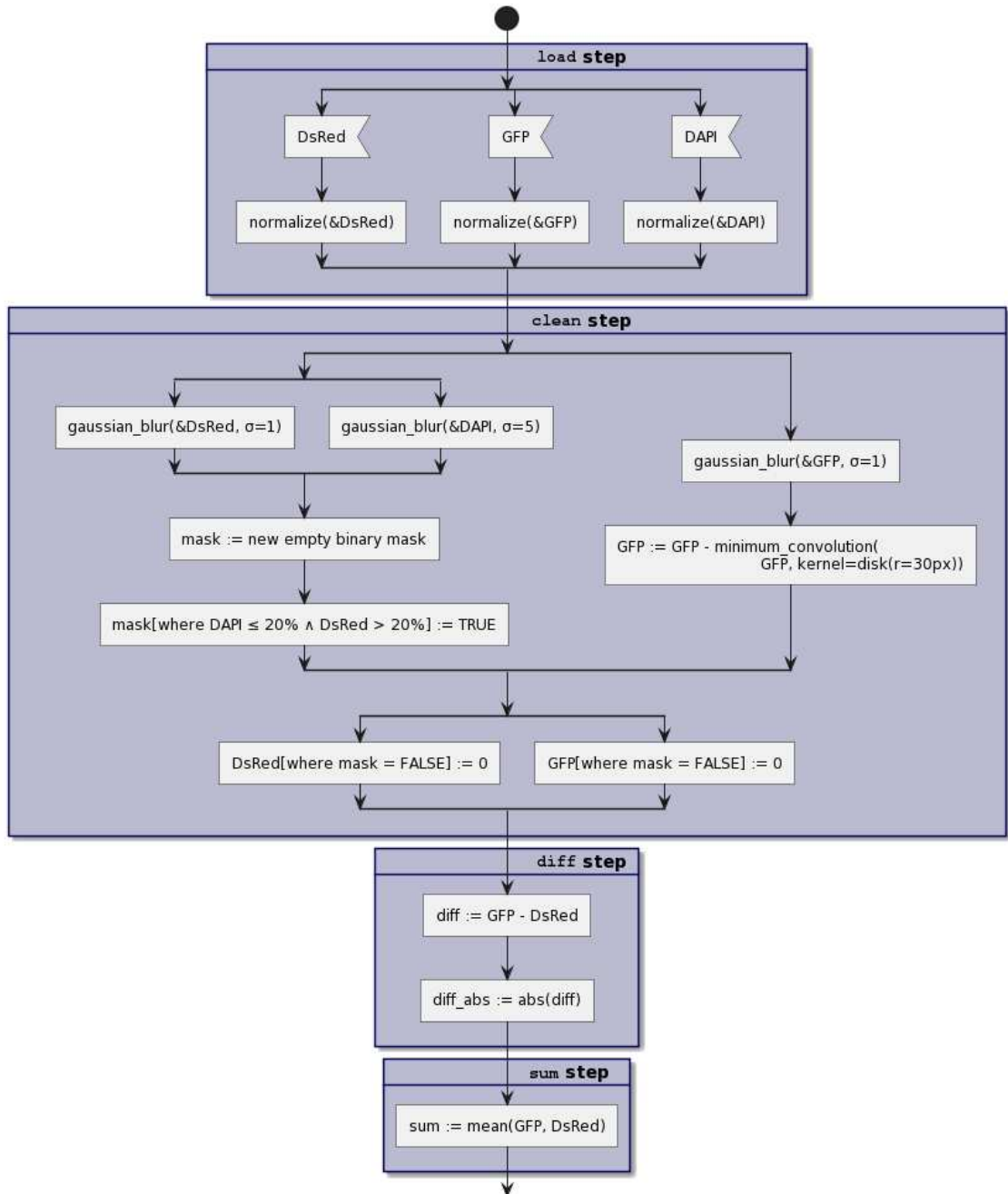
2.2. Algoritmus

Az algoritmus (*szakértői modell*) fontosabb lépéseit a 2.1 ábra mutatja meg. Ezeket alább, fejezetenként részletezem.

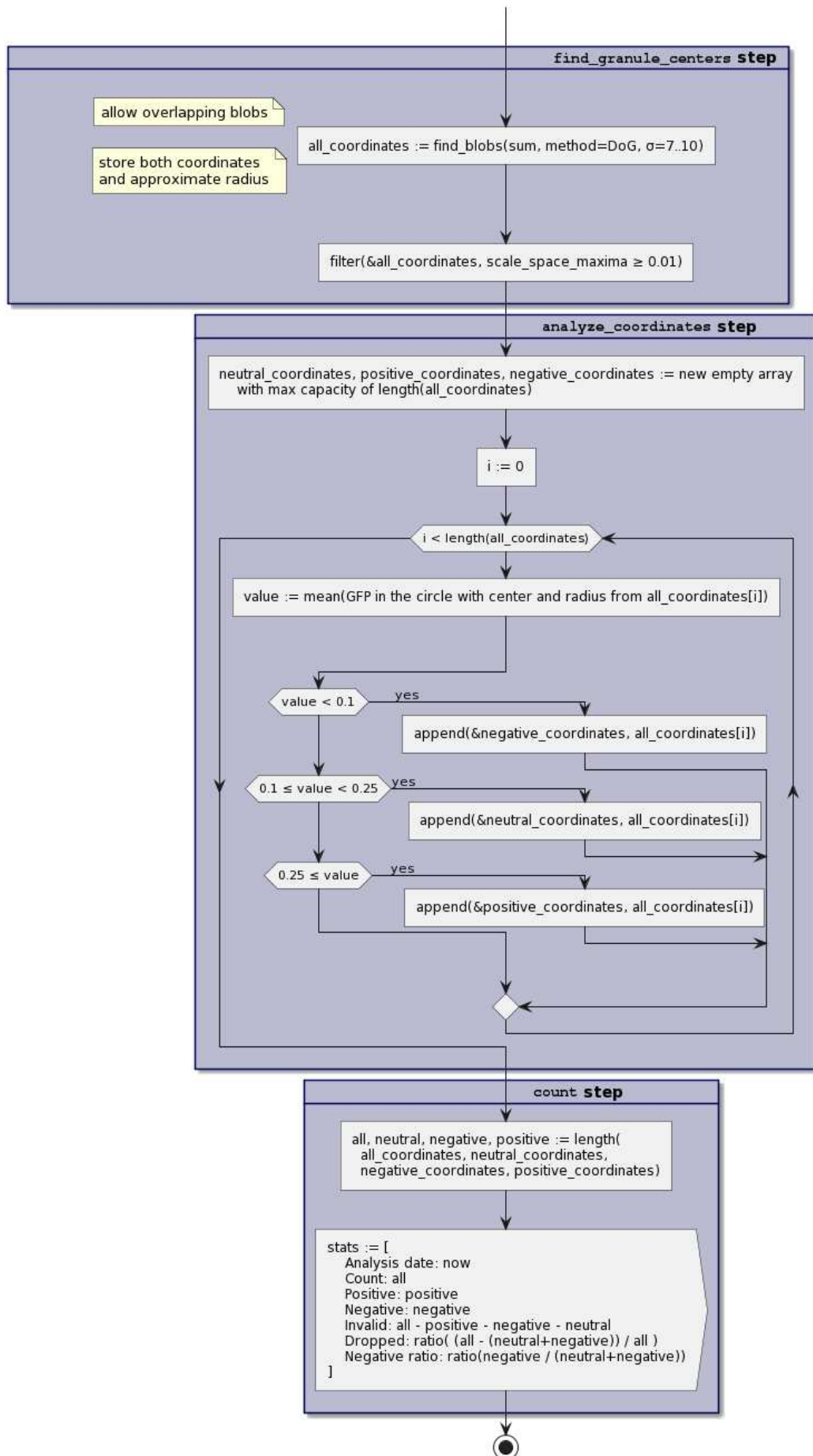
Az algoritmus helyes működése szempontjából fontosabb állításokat kiemeltem, ezekhez írtam feltételeket. Az algoritmus más kísérletre való adaptációja előtt mindenképpen szükséges a feltételeket végigolvasni és egyesével ellenőrizni, hogy az ott megfogalmazottak az adott kísérletre teljesülnek-e. Ha mind teljesül, az algoritmus átalakítás nélkül használható, ha nem, akkor a programkód átírásával (az algoritmus módosításával) az adott kísérletre kell szabni.

2.2.0. load

Mindenek előtt a forrásfelvételeket olvasom be a háttértárból. Egy apró technikai részlet, hogy az implementáció különböző fájlformátumokat támogat, így minden egyes csatornát külön, szürkeárnyaltos képként töltök be. Ezek a fájlformátumok a csatornákat



2.1. ábra. Összefoglaló folyamatábra a modellet megvalósító algoritmusról (1/2)



2.1. ábra. Összefoglaló folyamatábra a modellet megvalósító algoritmusról (2/2)

egy egyszerű sorszámmal azonosítják, így a képek paramétere, hogy melyik csatorna milyen információt hordoz, amit `dataset.toml` nevű fájlokban tárolok. Az információ átadásáért a `load` lépés a felelős.

A tárolt képek általában az információtartalom maximalizálása érdekében változó expozícióval készülnek el, így nem vesztek érdemi információt azzal, ha beolvasás után egyből normalizálom azokat (elosztom az összes intenzitás-értéket a maximális intenzitás-értékkel, így a maximális érték 1-re csökken az arányok megtartása mellett).

A beolvasott képeket DsRed, GFP és DAPI síkként (*plane*) tárolom el.

2.2.1. clean

Az első lépés (`clean`) feladata a zaj kiszűrése. Ez két al-lépésre bontható (ami a forráskódban nem jelenik meg).

Kezdeti zajsűrés

Sok képfeldolgozási algoritmus megbízhatatlanul működik részletes, zajos képen, a 2.2 ábra ezt mutatja be a limitáláson keresztül. Így a legelső művelet, amit a képeken végzek, egy struktúra csökkentés *Gauss-elmosás* (*gaussian blur*) formájában ($\sigma = 1$).

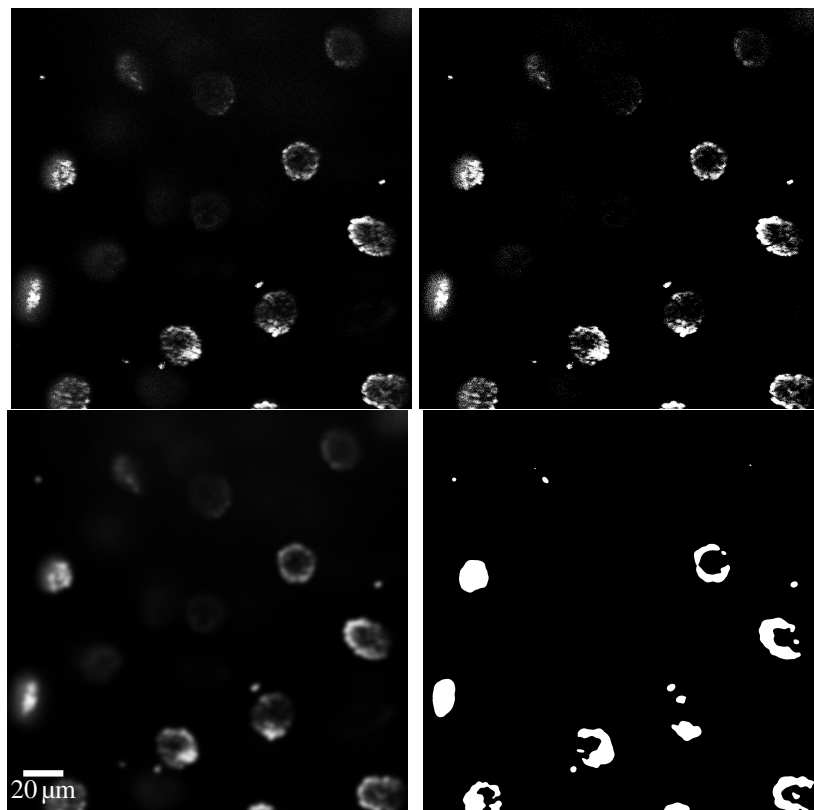
A modellemben feltételezem, hogy míg a DsRed *fluoreszcens marker* a vizsgálandó organellek kijelölésére használatos, a GFP *fluoreszcens marker* a jellemzők kijelölésére¹. Ezért a GFP jelnek egy speciális jellemzője lehet egy háttérjel, amelynek az erőssége más sejtes funkcióktól is függ, így a felvételeken régióról régióra változik. Ennek kiküszöbölése érdekében egy adaptív módszerrel normalizáltam az intenzitásokat: kísérletezésem alapján a lokális (30 pixel sugarú kör alakú környezetben) minimummal való csökkentés bizonyult a legjobb eredményt produkáló módszernek.

Maszkolás

A felvételeknek vannak olyan jól leírható részei, amelyekről ismert, hogy irreleváns, ezáltal megtévesztő információt tartalmaznak. A következőket vettem bele a modellbe:

- A sejtmagon belül az aktuális szeleten nincsenek granulumok, így az azon a területen lévő DsRed jel biztosan fókuszon kívülről érkezik

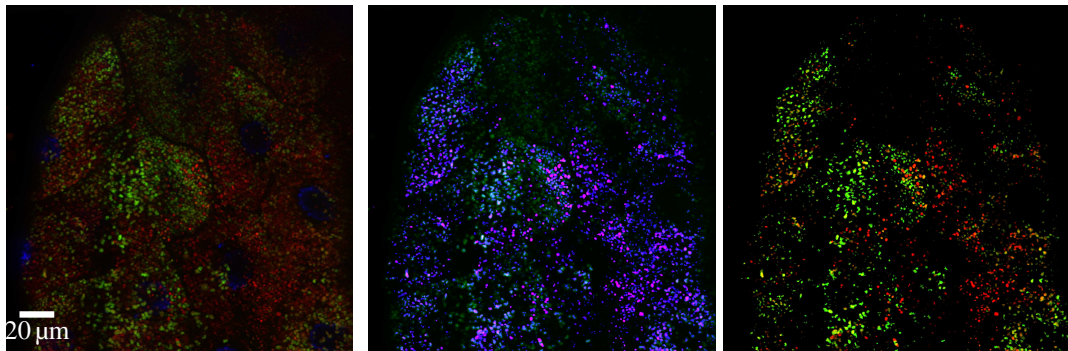
¹Fordított esetben a modell módosítása szükséges, a két csatorna megcserélése indokolt



2.2. ábra. DAPI *fluoreszcens marker* intenzitása egy képen, eredeti változatban (bal felső), közvetlenül limitálva (jobb felső) és *Gauss-elmosás* (*gaussian blur*) ($\sigma = 5$) után (bal alsó) illetve azután limitálva (jobb alsó). A maszkolás során a jobb alsóhoz változatot használok fel ebből a csatornából

- Feltételezve, hogy a GFP tetszőleges jellemzőket jelöl, a granulumok területén kívülről érkező jel irreleváns²

Ezeket a részeket figyelmen kívül kell hagyni. Ennek céljából létrehozok egy mask nevű, pixelenként egy logikai értéket tartalmazó síkot, ami a pozitív értékeket tartalmazza. Mivel a képeket normalizáltam, használhatok egyszerű limitálást. A mask síkon minden pixel értéke igaz akkor és csak akkor, ha a DAPI síkon az intenzitás nem nagyobb, mint 0.2, viszont a DsRed síkon nagyobb, mint 0.2.



2.3. ábra. Maszkolás. Az első a beolvasott (load) felvétel, második képen a maszkolás előtti DsRed (vörös), GFP (zöld) síkok és a generált mask sík (kék), a harmadik képen pedig maszkolás utáni DsRed (vörös) és GFP (zöld) látható. A felvétel tetején lévő sejt granulumjainak nagy része nincsen benne a fókusz síkban, így a lépés után az eltűnik.

Állítás. A mask sík hamis értékei alatt található pixelek figyelmen kívül hagyhatóak

Feltételek. Ezek kétféle területet takarhatnak:

Magas intenzitású DAPI (vagy *Gauss-elmosás (gaussian blur)* miatt annak környezete)

A magas DAPI emisszió fókuszban lévő sejtmag jelenlétére utal, így azon a területen nem lehetnek fókuszban lévő *szelekciós granulumok*.

Alacsony intenzitású DsRed (vagy *Gauss-elmosás (gaussian blur)* miatt annak környezete)

DsRed jelenlétének a hiánya cél-sejtorganellumok³ jelenlétének a hiányára utal. Vagy ott, vagy annak környezetében lehetnek csak vizsgálandó jelenségek.

A későbbi műveletek egyszerűsítése érdekében a DsRed és GFP síkokon az intenzitást 0 értékűre állítom ott, ahol nem teljesül a mask feltétele. A végeredményt a 2.3 ábra mutatja.

²A választott kísérlet esetében ideális esetben ilyen szituáció nem állhat fenn, de annak érdekében, hogy az algoritmust felkészíthessem más feladatokra, kiszűröm a potenciálisan problémát okozó jelet

³GlueFlux jelzőrendszer esetében ez *szelekciós granulum*

2.2.2. diff

A felvételeken a granulumok közül nem mind van fókuszban. Azonban szerencsére azok az objektumok, amelyek az egyik csatornán nincsenek fókuszban, a többi csatornán se lesznek. Így a jellemzőket jelölő *fluoreszcens marker* erősségét a vizsgálandó objektumokat kiemelő *fluoreszcens marker* erősségéhez érdemes viszonyítani.

Állítás. A granulumok területén belül a pixelenkénti intenzitások összegének eltolás utáni *trichotóm összehasonlítása* alkalmas osztályozásra

Feltételek. A *GlueFlux* jelzőrendszer esetében mind a két összehasonlítandó *fluoreszcens marker* ugyanoda lokalizálódik (a *szekekreációs granulumok* belsejébe), így annak a megállapítására, hogy elsavasodott-e az adott *szekekreációs granulum*, ideális módszer a *szekekreációs granulumok* területén belül a pixelenkénti intenzitások összegének összehasonlítása: Mennyivel több fotont (jelet) lehet savasodástól függetlenül is érzékelni, mint amit csak savasodás előtt?

Tapasztalatom szerint az összehasonlításra a különbség megbízhatóbb eredményt ad, mint az arány.

Mivel az implementációt Python nyelven írtam, és az optimalizáláshoz elsősorban numpy-t használok, érdemes minél több műveletet vektorizálni [33], ezért előzetesen elkészíték egy segédsíkot *diff* néven, ami a pixelenkénti intenzitáskülönbséget mutatja.

2.2.3. sum

Az egyes granulumok felismerése geometriai 3D-s gömbszerű struktúrák detektálásával történik. Alapvetően erre ideális megoldás lenne a DsRed jel figyelése, azonban bizonyos kísérletek esetében (például LampRing rendszer használata esetén [13]) a DsRed és a GFP egymástól különböző anyagokat jelöl, amelyek együttesen alkotják a *szekekreációs granulumon* belüli állományt.

Előzetesen elkészíték egy segédsíkot *sum* néven, amely a GFP és a DsRed pixelenkénti összegét tárolja, ezt fogom használni a későbbiekben a 3D-s gömbszerű struktúrák felismerésére.

Állítás. A *sum* sík alkalmas a 3D-s gömbszerű struktúrák felismerésére

Feltételek. A modell kidolgozása során feltételeztem, hogy a kísérlet két *fluoreszcens markert* alkalmaz (nevezzük őket GFP-nek és DsRed-nek), amelyek között kell kolokalizációt vizsgálni. Két eset lehetséges:

A GFP és a DsRed ugyanazt a fehérjét jelöli⁴ A GFP és a DsRed pontonként sztochasztikusan megegyező arányú minden egyedi granulumra, az összegzésük sztochasztikusan megfeleltethető egy egyszerű skalárral való szorzásnak

A GFP és a DsRed különböző fehérjét jelöl A GFP-vel és a DsRed-del megjelölt fehérjék metszete üres halmaz. A modellben feltételezem, hogy az *uniójuk* a granulumon belüli állomány. Így a pixelenkénti összegük éppen kiadja a granulumon belüli állományt.

2.2.4. find granule centers

Az egyes granulumok felismerésére az *DoG*-piramison alapuló *sarokdetektálás* (*blob detection*) algoritmust [25] alkalmazom, 7-10 között mozgó σ értékkel. A *DoG* alapú algoritmus előnye az, hogy a futásideje és a pontossága ideális egyensúlyban van, szemben a *DoH* és *LoG* alapú algoritmusokkal. Ezekről részletesebben a mellékelt A.4.1 részben írok.

Az algoritmus akár a legkisebb zajt is foltként azonosíthatja. Kísérletezés útján megállapítottam, hogy *GlueFlux* jelzőrendszer esetében az ideális szűrőfeltétel valahol 0.01 intenzitású *scale-space*-maxima környékén van: Ez az az érték, amivel a granulumfelismerési hibák mennyisége minimalizálható⁵. A 2.4 ábra mutatja a szűrés eredményét, ami csak a fókuszban lévő granulumokat tartja meg.

Állítás. A Lowe féle *DoG* alapú algoritmus [25] a $\max_{l \in \text{scale-space}} l[x, y] \geq 0.01$ szűrési feltétellel megfelelő a granulumok beazonosítására

⁴A *GlueFlux* jelzőrendszer esetében a GFP és a DsRed ugyanazt a fehérjét jelöli

⁵Kísérletezésem során a hibátlan eredményre törekedtem. Tapasztalatom szerint az algoritmus tökéletesen működik (több ezer dokumentálatlan esetet kézzel megvizsgálva egyben sem találtam hibát, dokumentált vakteszt esetén (3. fejezet) minden tekintetben kevesebb, mint 5% hibát találtam, azonban ezek többnyire manuálisan is nehezen osztályozhatóak), de ezen a ponton az algoritmusom indoklási láncá sérül, így alkalmazás esetén fontos az egyes felvételek kvantifikációjának szemrevételezése a *dashboard* segítségével

Feltételek. A visualize coordinates lépés all síkja (vagy a *dashboard*dal is betölthető⁶, stats.csv fájlba elmentett pontok listája) nem tartalmaz se hamis pozitívokat, se hamis negatívokat (2.4 ábra). Ez valószínűsíthető (3.1.1 Validálás szakértői kvantifikáció alapján), de szemrevételezéssel ellenőrizendő az állítás belátásához.

2.2.5. analyze coordinates

A granulumok osztályozásáért az analyze coordinates lépés felel. Ez a *szakértői modell* leginkább kísérlet-specifikus része, így nem *GlueFlux* jelzőrendszer esetén minden bizonnyal ennek a dolgozatnak a keretein túlmutató implementáció szükséges.

Állítás. A analyze coordinates lépés megfelelő

Feltételek. A kísérlet során a Csizmadia és tsai. által is használt *GlueFlux* jelzőrendszer⁷ van használatban

A 2.2.2. állítás szerinti eltolás az algoritmusom fő *hiperparamétere*. Két jellemző hiperparamétert rögzítettem az implementációban, ezek szerint a granulumokat három osztályba sorolom be:

Pozitív ha a különbség a nagyobb hiperparaméter fölötti

Semleges ha a különbség a két hiperparaméter közé esik

Negatív ha a különbség a kisebb hiperparaméter alatti

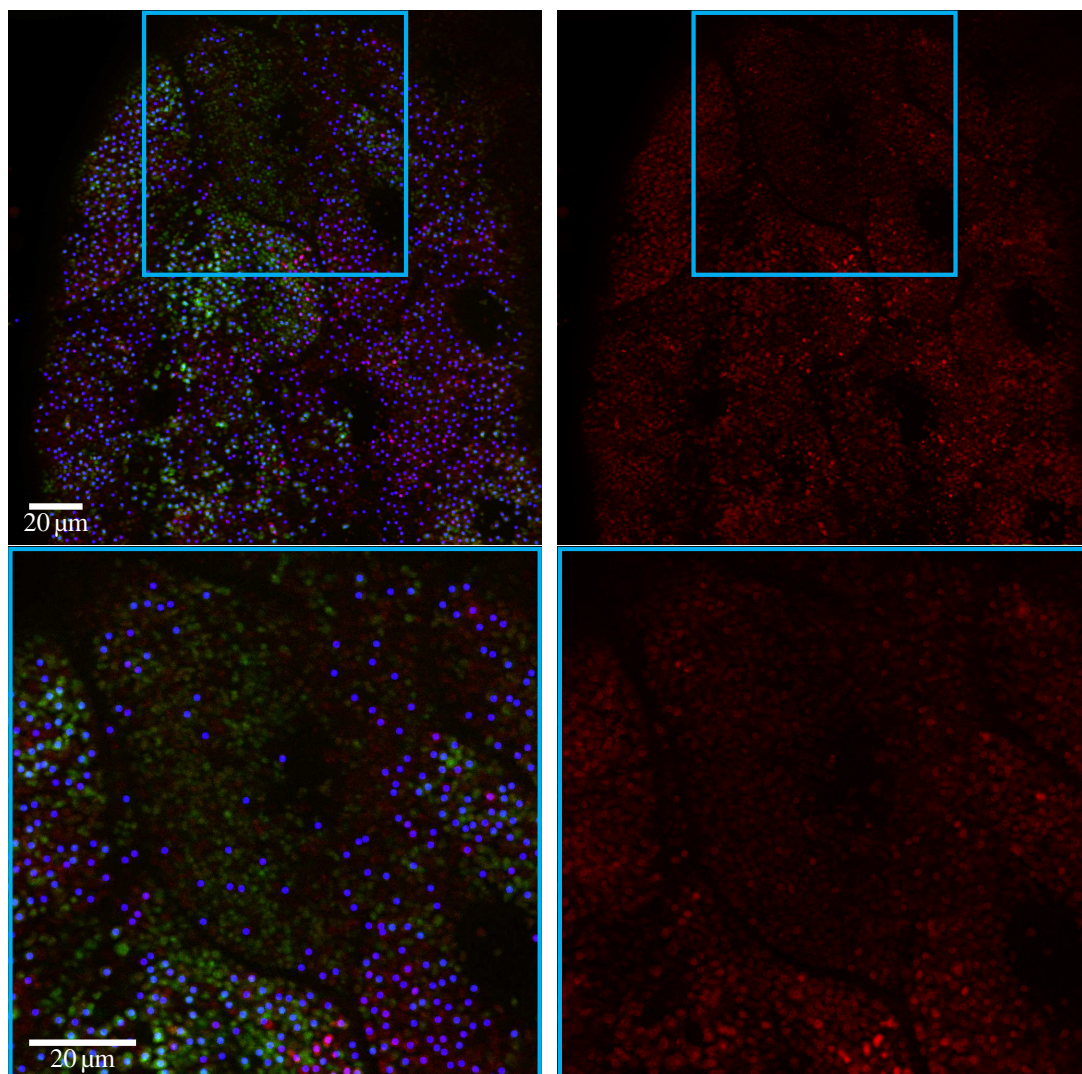
Az adott kísérlet érzékenysége szerint a semleges találatok értelmezhetőek semlegesként, pozitívként és negatívként is.

2.2.6. count

Az utolsó lépése a modellnek az összegzés (count). Ennek során az összes, a pozitív, a negatív és egyértelműen téves találatok számát (pár tájékoztató jellegű metaadattal együtt) szöveges formában a képernyőre írom, illetve opcionálisan fájlba mentem.

⁶A *dashboard* rendelkezik egy olyan funkcióval, mellyel könnyedén, egérrel bekarikázva ki lehet zárni a tévesen felismert granulumokat

⁷Tamás Csizmadia és tsai. "Molecular mechanisms of developmentally programmed crinophagy in *Drosophila*". *Journal of Cell Biology* 217.1 (2017), 361–374. old. DOI: 10.1083/jcb.201702145.



2.4. ábra. Felismert granulumok vizualizációja. Az első képen a vörös illetve zöld csatornán a betöltött nyers adatok láthatóak, kék színnel pedig a felismert granulumok vannak jelölve (a11 sík); a második képen a vörös csatornát emelem ki; az alsó sorban a felső képek tetejének (cián keret) kinagyított változata látható. Ideális esetben kék jelölő éppen kvantifikálásra alkalmas granulumokra van elhelyezve. A második képről leolvasható, hogy a felső-középső részen sok *szekréciós granulum* nincsen fókuszban, így csak 1-2 kék jelölőt látunk azon a régióban, azonban közelről megnézve az elhelyezett jelölők éppen értékelhető granulumokat takarnak.

A 2.3 forráskódról leolvasható, hogy az algoritmus összesen 1664 objektumot talált, amiből 297 egyértelműen *pozitív*, 878 pedig egyértelműen *negatív* minta.

```
1 Analysis date: 2022-11-28 18:11:25.496636
2 Count: 1664
3 Positive: 297
4 Negative: 876
5 Neutral: 484
6 Invalid: 7
7 Dropped: 0.42%
8 Negative ratio: 52.87%
```

2.3. forráskód. A modell által generált összegzés

3. fejezet

Az eredmények értékelése

3.1. Statisztikai validálás

A modell statisztikai validálásához a konvencionális 5%-os szignifikancia-szintet használok. Kétféle *ground truth* adathalmazzal végzek kísérletet. Az egyik esetben szakértői kvantifikációt használok, a másik esetben jól felderített biológiai funkciókra támaszkodok.

3.1.1. Validálás szakértői kvantifikáció alapján

Néhány, reprezentatívan¹ választott felvétel granulumjait kézzel átnéztem, és összevetettem az algoritmus eredményével (3.1 táblázat).

A hiba aránya² egyik esetben sem ért el 5%-os értéket, annak ellenére, hogy a lehető legszigorúbb számítást alkalmaztam. (Valójában a statisztika szabadsági foka ehhez képest nagyban lecsökken, mivel a végső arányban csak a hibák különbsége jelenik meg. Továbbá érdemes különválasztani a téves *pozitív* és a téves *negatív* eseteket, így ez a szám még két rész összegére bontható.) Az alá- és fölébecsült esetek aránya még kisebb. A kísérlet során mind a közel 1309 pontot egyedileg, a többitől függetlenül vizsgáltam meg.

Látható, hogy az összesített eltérés elenyésző. Sajnos a *negatív* hibának enyhén nagyobb az esélye, de ez a szakértő szigorúságán is múlik (honnantól tekintünk egy

¹Igyekeztem a felvételen szereplő állatok relatív életkora, az összes granulumszám és a pozitivitás dimenzióknak mindegyik végletét érintő felvételhalmazt összeállítani kontrollfelvételekből, egyéb tekintetben véletlenszerűen válogattam

²Hiba% = $\frac{\text{a rendszer által figyelmen kívül hagyott granulumok} + \text{a rendszer által tévesen granulumnak tekintett objektumok}}{\text{felvételen látható összes granulum}}$

Belső azonosító Pozitivitás ^[8] Össz. ^[9]	Alá ^[1]	Fölé ^[2]	Hiba _{poz} ^{[3],[5]}	Hiba _{neg} ^{[4],[5]}	Hiba% ^[10]
GF_fos_RNAi_-1h-0108 35.2% 466	0 ^[6]	1 ^[6]	+2 – 2	+9 – 10	4.9%
Rabex5_RNAi_GF_0h-0010 12.7% 569	2 ^[6]	0 ^[6]	+1 – 1	+2 – 3	1.2%
Rabex5_RNAi_GF_0h-0004 6.3% 270	1 ^[6]	0 ^{[6],[7]}	+2 – 0	+2 – 4	3.0%

3.1. táblázat. Kézzel készített hibastatisztika. Megjegyzések: ^[1]Olyan granulumok száma, amelyek alacsonyabb rendű osztályba kerültek, mint kellett volna: semleges helyett *negatív*, vagy *pozitív* helyett semleges, ^[2]Olyan granulumok száma, amelyek magasabb rendű osztályba kerültek, mint kellett volna: *negatív* helyett semleges, vagy semleges helyett *pozitív*, ^[3]Eltérés olyan granulumok számában, amelyeket a rendszer nem ismert fel, de fókuszban vannak és *pozitív*nak tekintendők, ^[4]Eltérés olyan granulumok számában, amelyeket a rendszer nem ismert fel, de fókuszban vannak és *negatív*nak tekintendők, ^[5] $+a - b$ jelentése: a másodfajú és b elsőfajú hiba, ^[6]Nem volt olyan granulum, ahol dupla osztályugrás történt volna, ^[7]Ezen a felvételen a nagy háttérzaj miatt 41 granulum esetén szemrevételezéssel megállapíthatatlan volt, hogy milyen osztályba kell azokat besorolni, így ezeket a statisztikába nem számolom bele, ^[8]A rendszer által számolt pozitivitási arány, ^[9]Összes olyan granulum száma, amelyet a rendszer felismert, ^[10]Az összes hiba összege és a granulumok teljes számának aránya

granulumot kettőnek). Szerencsére ezek az esetek mind kiszűrhetőek az eredmény gyors (5–10 másodperces) szemrevételezésével, ami biológiai funkciók kizárása miatt is feltétlenül szükséges szakszerű adatfeldolgozás során.

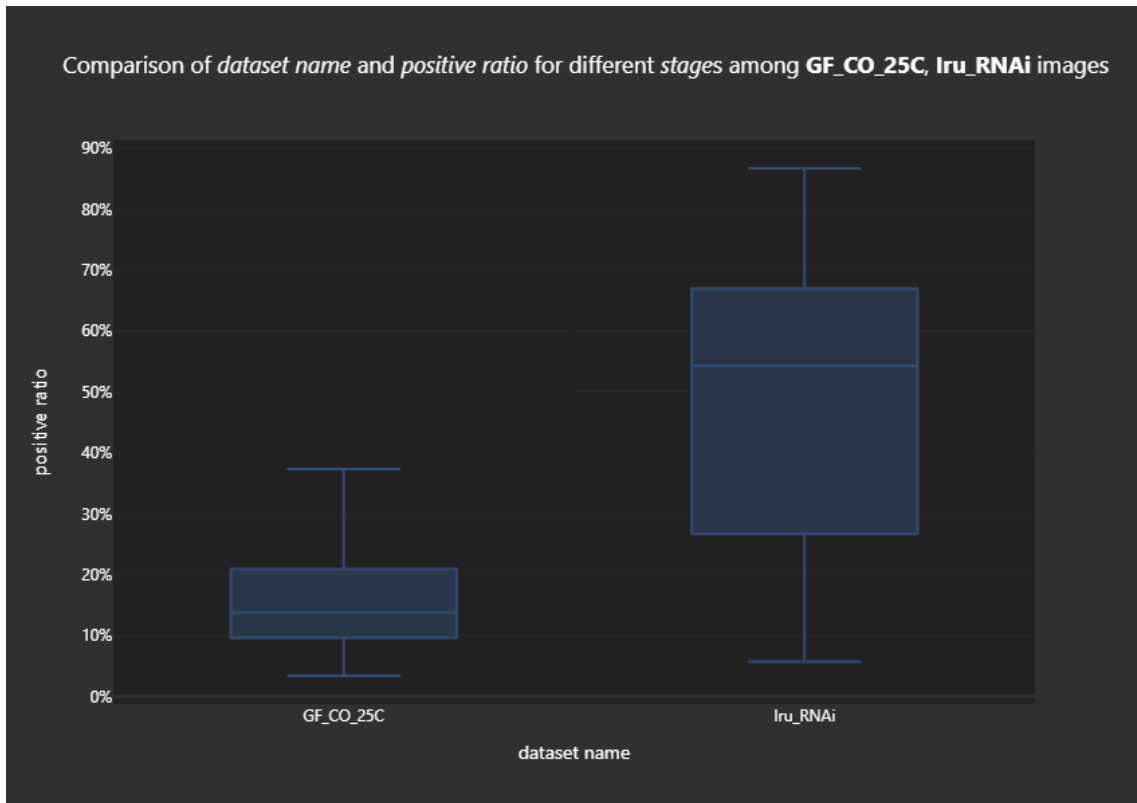
A fentiek tekintetében nagyobb hangsúlyt fektetek a gyakorlati feladatokkal történő validálásra.

3.1.2. Validálás biológiai funkciók segítségével

A modellem validálását elvégeztem olyan biológiai kísérleteken, amelyeket az *ELTE Biológiai intézet Anatómiai Sejt- és Fejlődésbiológiai Tanszék (ELTE ASF)* laborjaiban kontrollkísérletként végeztek, így a kimenetelük felhasználható, mint *ground truth*.

A savasodás hőmérsékletfüggése

A jelenlegi modell egyik validálását és egyben első komolyabb sikertörténetét egy váratlan esetnek köszönhetem. Az ELTE látványosi kampuszának déli épületében 2022



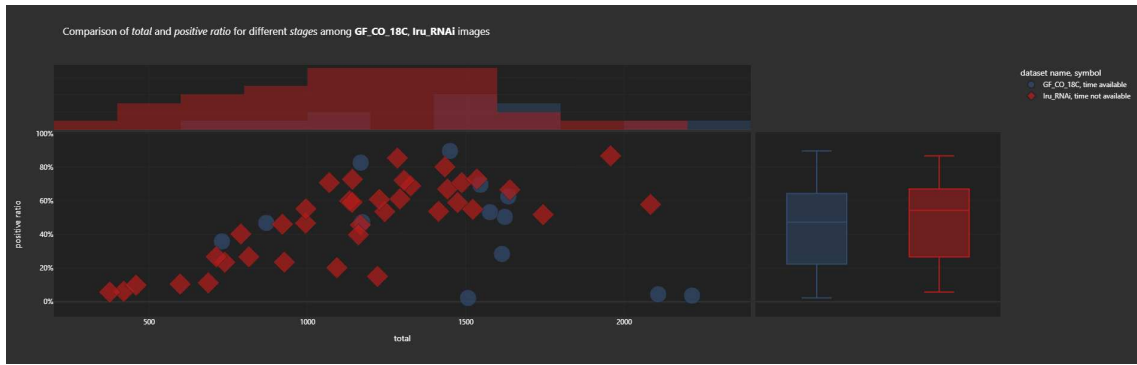
3.1. ábra. Kontroll és megismételt géncsendesítés (Iru^{RNAi}) összehasonlítása, y tengelyen a pozitív granulumok arányával az összeshez viszonyítva. Amennyiben az elképzelések helyesek, mind a két minta pozitivitása alacsony.

telén egy időre lecsökkentették a beltéri hőmérsékletet 18 °C környékére. Egy korábbi, *GlueFlux* jelzőrendszert használó kísérletet (ami arra engedett következtetni, hogy az Iru^{RNAi} géncsendesítés nem befolyásolja az elsavasodást, tehát kevés pozitív granulumot eredményez) megismételték, mivel régebben kevés felvétel készült. A várakozásokkal ellentétben meglepő módon azt tapasztalták, hogy a korábbi kontrollcsoporthoz képest jóval alacsonyabb fokú volt az elsavasodás (tehát sok volt a pozitív granulum).

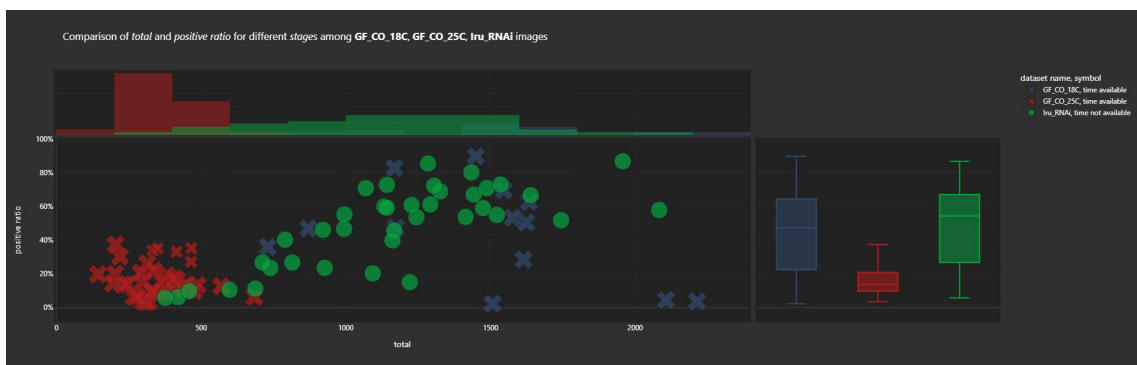
Az eredmények a 3.1 ábrán láthatóak. A grafikont megnézve szembetűnő, hogy a géncsendesített törzs nem hozza az alacsony pozitivitást.

Szerencsére ennek ellenére a felvételeket elmentették és megengedték, hogy azokra lefuttassam a programomat. Az eredményeket a 3.2 ábra mutatja, meglepő módon nincs szembetűnő különbség a két adathalmaz között.

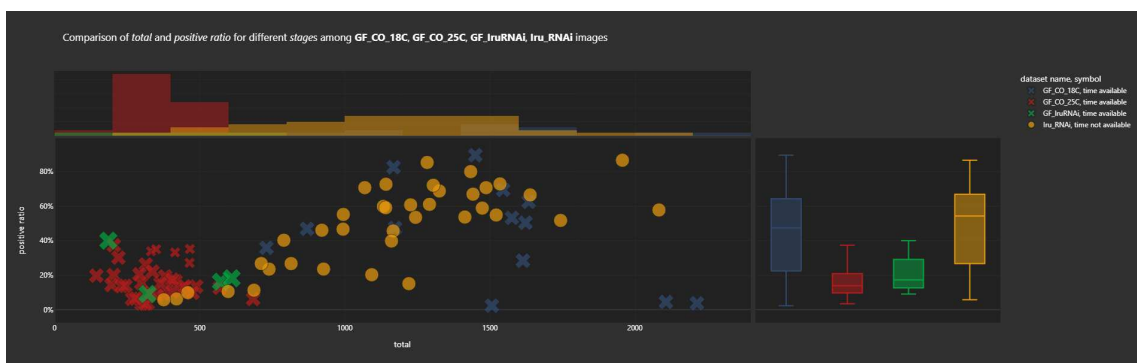
Feltűnt, hogy a pozitivitás szokatlanul magas, így elkértem a korábbi kísérletek kontrollfelvételeinek gyűjteményét, és azokra is lefuttattam a programot, azonban meglepő eredményt tapasztaltam: Jelentős eltérés van a megismételt kísérlet kontrollja és a korábbi kísérletek kontrolljai között (3.3 ábra).



3.2. ábra. Megismételt kísérlet kontroll (kék) és géncsenedesített (*Iru^{RNAi}*) (piros) felvételeinek összehasonlítása, y tengelyen a *pozitív* granulumok aránya, x tengelyen az összes granulum száma az adott felvételen.



3.3. ábra. Géncsenedesített (*Iru^{RNAi}*) és különböző kontrollfelvételek összehasonlítása, y tengelyen a *pozitív* granulumok arányával az összeshez (x tengely) viszonyítva. Kék színnel a 18°C-os környezetben végrehajtott kontrollkísérlet, piros színnel a 25°C-os környezetben végrehajtott összes korábbi kontrollkísérlet közül a közeli felvételek, zöld színnel pedig a géncsenedesített törzs felvételei. Az új felvételeken (kék és zöld) a granulumok magas számának a magyarázata az, hogy a gyors közbelépésemnek köszönhetően nagyobb területet felölelő felvételek kerültek mentésre.



3.4. ábra. Különböző hőmérsékleten végzett géncsenedesített (*Iru^{RNAi}*) és kontroll felvételek összehasonlítása, y tengelyen a *pozitív* granulumok arányával az összeshez (x tengely) viszonyítva. Piros és zöld színnel a 25°C-on végzett kontroll és kísérleti csoport eredményei, míg kék és sárga színnel a 18°C-on végzett kontroll és kísérleti csoport eredményei.

Sok esetet szűrőpróbaszerűen és a kifejezetten hibagyánús régiókat is átnézve szemrevételezéssel megvizsgáltam, és nem találtam tévesen felismert granulomot. Azonban ha az eredeti pár felvételt is ráhelyezzük a grafikonra, világosan látszik, hogy a 18°C-os kontroll és a 18°C-os kísérleti csoport illetve a 25°C-os kontroll és a 25°C-os kísérleti csoport páronként hasonló eredményeket produkál (3.4 ábra).

Időközben a kísérlet felelőisével közösen rájöttünk, hogy a *GlueFlux* jelzőrendszer alapvető feltétele, a lizoszómák elsavasodása 17.5°C-on jelentősen kisebb mértékben megy végbe, mint 18°C-on és fölötte [40].

A fenti bekezdés alapján a történet a modellem használhatóságának egy vaktesztje. Habár ez a modellem validációja szempontjából nem releváns, a 18°C-os mintákon az átlagos pozitivitás szignifikánsan nagyobb, mint a 25°C-os mintákon³.

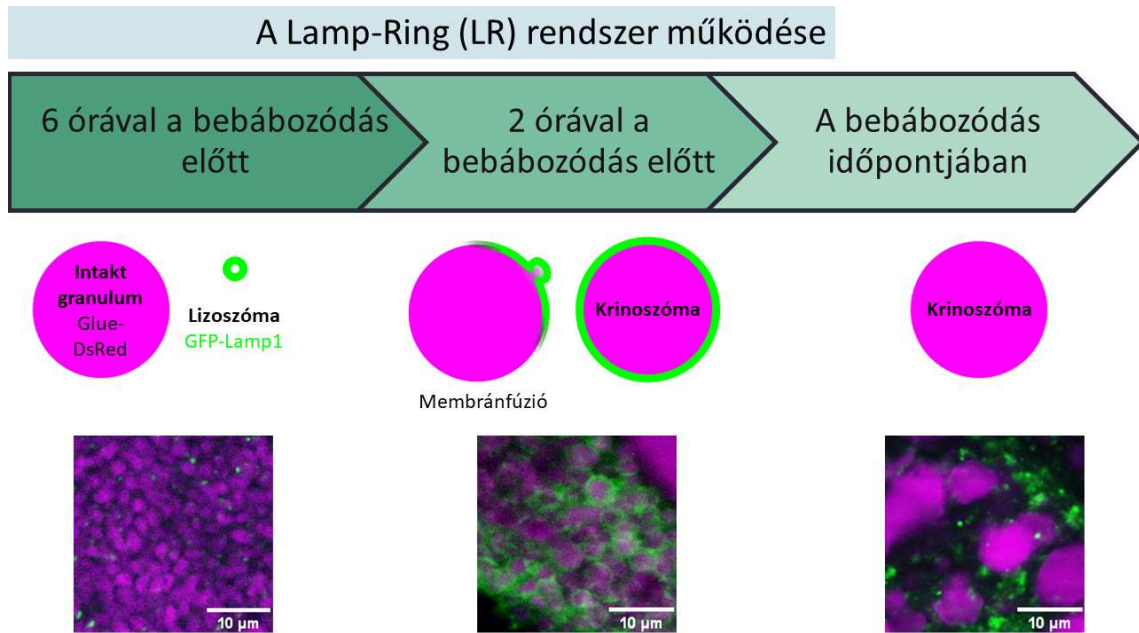
Korábbi kísérlet

A modellem megbízható validálása érdekében végrehajtottam egy sokkal kézzelfoghatóbb tesztet is.

A *GlueFlux* jelzőrendszer működéséből adódik, hogy minden sejtben, így minden nyálmirigyben is az idő előrehaladtával a *pozitív* granulomok száma monoton csökken. Sajnos a granulomok valós idejű megfigyelése jelenleg meglehetősen költséges és körülményes [11], de szerencsére a felvétel elkészítése előtt felismerhető a használt állatok bábozódáshoz viszonyított életkora (*time Relative to Puparium Formation (RPF)*). Több kontrollfelvételhez rendelkezésemre állt az *RPF* érték, így a modellemnek egy jó ellenőrzése, hogy tetszőleges célból készült kontrollfelvételen ki tudja-e mutatni a pozitivitás csökkenését.

Ennek az ellenőrzésére egy egyoldali Student féle T-tesztet végeztem, összehasonlítva a -1 *RPF* és 0 *RPF* értékű kontrollfelvételeket. A teszt alapján a p-érték 0.7%, így elvethető az, hogy a program nem mutatna ki pozitivitás csökkenést az időben eltérő felvételek között, ezáltal valószínűsíthető, hogy az algoritmusom helyesen működik.

³A `t.test(stats$positive.ratio[stats$dataset.name=='GF_CO_18C'], stats$positive.ratio[stats$dataset.name=='GF_CO_25C'], alternative = "greater")` R parancs segítségével megállapítható, hogy a p-érték 0.03



3.5. ábra. A LampRing rendszer működése. A középső sorban sematikus ábrákkal, az alsó sorban pedig mikroszkópos felvételeken követhető nyomon a gyűrűk kialakulásának folyamata LampRing rendszer esetén [9]. A magenta szín DsRed jelre, a zöld pedig GFP jelre utal.

3.2. Sikertelen kísérletek

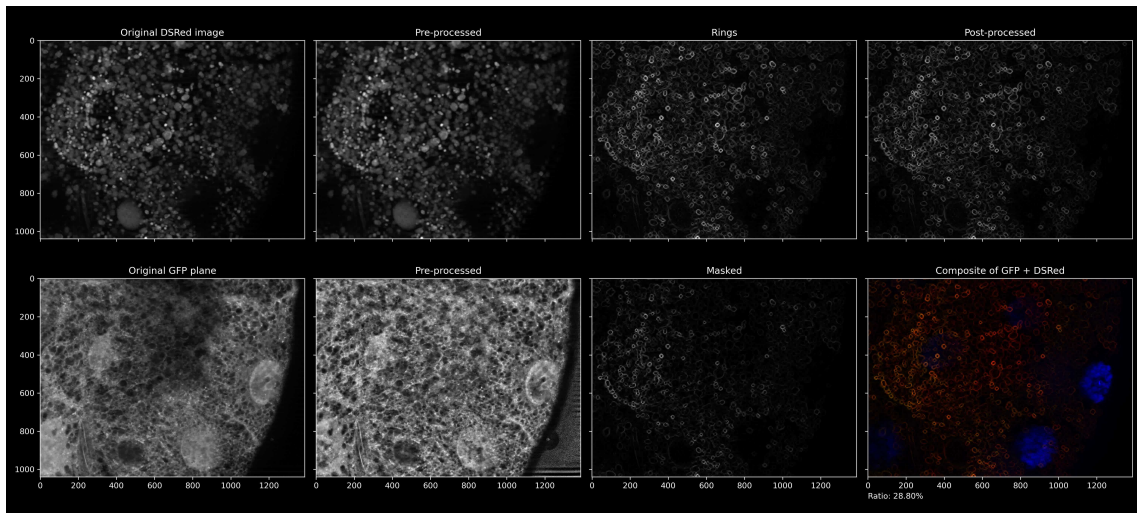
Egy másik, nagyon hasonló biológiai probléma a granulumok membránjaira lokalizálódó GFP fehérjék (gyűrűk) kvantifikálása. Ez a LampRing rendszer [13] alkalmazása esetén merül fel feladatként. A folyamat a 3.5 ábrán látható.

Nagyon sokat próbálkoztam morfológiai megközelítésű gyűrűkeresési algoritmus kidolgozásával is.

Ezeket a dolgotban nem részletezem olyan mélységgel, mint a működő és használatra javasolt algoritmust, mivel ezeket még tovább szeretném fejleszteni, működő állapotba hozni, azonban megemlítem, mint a dolgozat keretében végzett munkákat.

Pixelenkénti összesítés

A legelső implementációt sztochasztikus alapokra helyeztem. Ennek során minden egyes pixelre kiszámítottam két értéket: Mekkora eséllyel része membránnak DsRed jelek alapján és mekkora eséllyel pozitív GFP jelek alapján membránon. Ezek után a valószínűségeket egyszerű összeadással az egész képre összegeztem, és az alapján számítottam arányt. A működését a 3.6 ábra mutatja.



3.6. ábra. Az eredeti implementáció működése. A konklúziót az utolsó panelről lehet levonni, amelyen vörös színnel a membránok, zöld színnel pedig a gyűrűk láthatóak, a két szín keverékéből a színárnyalat alapján meg lehet mondani, hogy a narancssárga az pozitív, GFP gyűrűs granulumot, míg a vörös az negatív, GFP gyűrű nélküli granulumot jelöl.

A DsRed jelet dilatáció és erózió különbségével membránokká alakítottam, majd pixelenként összeszoroztam a GFP jellel, így a GFP jelnek csak azok a részei maradtak erősek, ahol valóban a membránon vagyunk, így a probléma egy kolokalizációs jellegű feladattá válik [23].

Ennek a megközelítésnek az előnye a (jól paraméterezhetőség mellett) skalár jellegű eredmény. Minden granulumról tudok adni $(0, 1)$ nyílt tartományból származó eredményt, ami leírja, hogy mennyire jó minőségű az adott gyűrű, így ezek összege az egész kép minőségét jellemzi. Hátránya viszont a nehéz szakértői ellenőrzés. A szakértőktől sokkal könnyebben begyűjthető kategorikus adat mint valószínűségi adat.

Ennek a megközelítésnek egy továbbfejlesztési iránya mesterségesen kiválasztott és — akár 50×50 pixel méretűre — kivágott mintákon való validálás és/vagy kalibrálás lehet, illetve érdemes kipróbálni más jellegű (pl. négydimenziós) felvételek használatával is [41].

Hough transzformáció

Az előző algoritmus ötletét folytatva a karikákat (mesterséges gyűrűket) ki lehet listázni Hough gyűrűkereső algoritmusával [42] segítségével.

Ennek a módszernek a hátránya a pontatlanság volt, amelynek az az oka, hogy a

granulumok alakja olyan messze van a szabályos körtől, hogy a szűrőfeltételek szigorítása a zaj kiszűrése mellett a granulumok nagy részét is kiszűri. Azonban mint ahogy az 1.4 *Célkitűzések* alfejezetben is megfogalmaztam, az egyik legfontosabb követelmény, hogy a minta reprezentatív legyen, így ezt a módszert mindenképp el kell vetni.

Ennek a megközelítésnek egy továbbfejlesztési iránya jobb gyűrűkereső algoritmusok [43] használata lehet.

Flood

Megpróbálkoztam intuitív alapon *flood-fill algoritmus* alkalmazásával. Ennek a megközelítésnek a legnagyobb kihívása talán a kontúr kijelölése, mivel több granulum akár átfedésben is lehet amellest, hogy különböző intenzitással jelennek meg a képen.

A kihívás leküzdésére megpróbálkoztam egy módosított implementációval, amiben minden pixel esetén a középponttól való távolság alapján számítok limit-értéket a kontúr megállapításához.

Egyre több jellemző figyelembe vételével az algoritmus komplexitása folyamatosan nő, így úgy vélem, ilyen jellegű probléma esetén nagyobb haszna volna egy neurális háló alapú implementációnak.

Ennek a megközelítésnek egy továbbfejlesztési iránya természetes jelleggel adódik, a flood fill esetén a feldolgozás sorrendjének és a kontúr meghatározásának javításával.

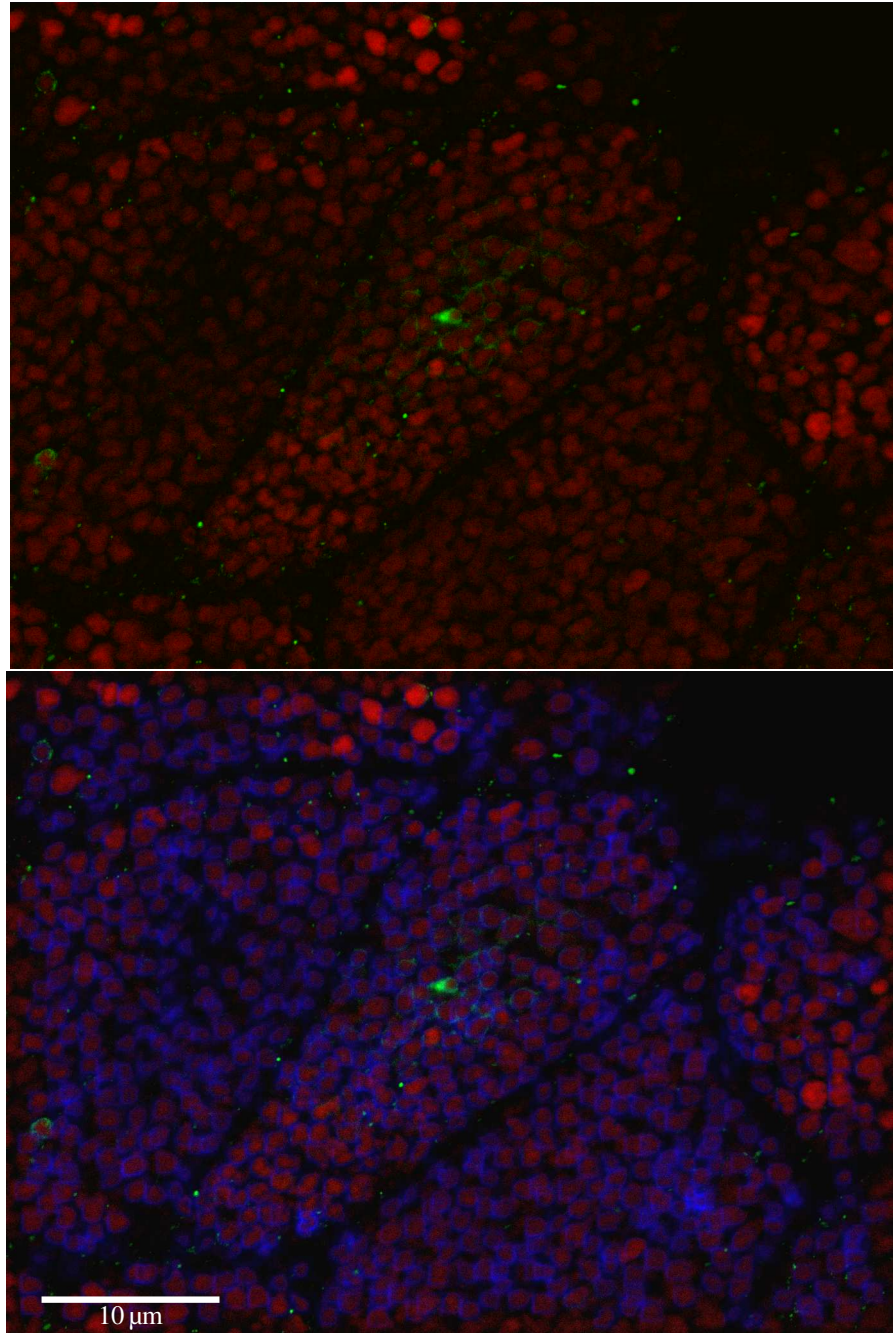
Átmérő meghatározása DoG piramis használatával

A Lowe féle foltkereső algoritmus [25] egy közelítő átmérőt is meghatároz. Ez kiindulópontja lehet a lehetséges gyűrűk körrel való közelítésére. Azonban egyrészt sajnos a granulumok membránja nem feltétlenül követi a kört, ezért gyanús, hogy nem teljesítené a megközelítési mód a hiba reprezentativitására való követelményt, másrészt Lowe algoritmus [25] a foltok átmérőinek felismeréséhez nagyon pontatlan közelítést ad, abból adódóan hogy a kifejlesztése során nem valódi foltok, hanem sarokpontok detektálására koncentráltak.

Ennek a megközelítésnek egy továbbfejlesztési iránya alternatív algoritmus keresését és a módszer erős igazolását (akár statisztikai módszerekkel, akár indoklással) jelentené.

Mélytanulás használata

Mélytanuló algoritmus bevetésével meggyőző eredményeket értem el, azonban még nem sikerült olyan neurális hálót találnom, amely teljesíti a 1.4 Célkitűzések alfejezetben támasztott elvárásokat, mely szerint az algoritmusnak átláthatónak kell lennie. A tervem egy egyszerűbb hálózattal való kísérletezés. A legjobb kísérleteim egyik demonstrációs jellegű eredménye a 3.7 ábrán látható.



3.7. ábra. A vörös és a zöld csatornán a szokásos GFP és DsRed látható, a kék csatornán azonban *MI* segítségével mesterségesen előállított gyűrűk

4. fejezet

További kutatási lehetőségek

A kutatást két irányban érdemes folytatni: Egyrészt a jelenlegi modellt és algoritmust lehet még tökéletesíteni, másrészt a jelenlegi munkám jó alapul szolgál más, hasonló feladatok kvantifikálásához.

Jelenlegi modell tökéletesítése

Egyrészt a jelenlegi implementáció egy proof-of-concept-szerű program, azaz a futási idő minimalizálása nem tartozott a céljaim közé. Alapos átgondolással és esetleg másik programozási nyelvre való átírással jelentős teljesítménybeli javulást lehet elérni, de már a többszálúsítása is nagyon hasznos lenne. Ideális esetben akár az egyes felvételek feldolgozását több különböző, hálózatban összekapcsolt számítógép is végezhetné. Lokálisan gondolkodva videokártya használatba vétele is egy jó irány az optimalizálásához.

Ezen kívül a dolgozatban a *find granule centers* lépés egy szürke folt, nem teljesen körültekintő a lépés indoklása. Egy lehetséges kutatási irány a lépés igazolása, illetve a Lowe által kidolgozott *DoG* alapú algoritmus [25] helyett jobb algoritmusok keresése a 2.2.4 szakaszban taglaltak szerint, egy másik lehetséges kutatási irány pedig 3D-s (a biológiai szakirodalomban *z-slice*-nak nevezett) felvételek felhasználása a lokalizálás pontosításához, amely felhasználható lenne a 2D-s eljárás validálásához, mivel a mintavételezés felbontása nagyságrendekkel nagyobb, mint 2D-ban [44].

Más feladatok kvantifikálása

A dolgozatomban remek alapként szolgál más hasonló munkák számára. Egyrészt tömören összefoglaltam a *krinofágia* kutatásával kapcsolatos olyan tudnivalókat, amik egy informatikus számára relevánsak lehetnek, másrészt összefoglaltam a morfológiai jellegű képfeldolgozás alapvető algoritmusait, amelyek mikroszkópos képek elemzése esetén hasznosak lehetnek.

Ezen kívül a kikísérletezett algoritmusomat részben vagy egészében fel lehet használni más jellegű mikroszkópos képek elemzésére. Használható egyrészt közvetlenül neurális háló augmentációjaként (amennyiben a felvételek hasonló méretskálában mozognak), másrészt amennyiben a tételek formájában megfogalmazott feltételek közül nem sérül mindegyik, az algoritmusom egy-egy lépése akár átalakítás nélkül felhasználható és/vagy a programnak csak kis részén kell módosítani. Ilyen például a membránokra lokalizálódó gyűrűszerű struktúrák keresése.

Harmadrészt a járulékosan született szoftver önmagában is kiváló alapként szolgál mikroszkópos képek feldolgozására. A *pipeline_lib.py* könyvtárral könnyedén építhetők hasonló felépítésű, de teljesen más megközelítésű algoritmusok, amikben a hiba kereséséhez a felhasználói felületem nagyobb segítséget nyújt az ismert integrált programozói környezeteknél (IDE-knél). Ha az alternatív feladat hasonlóan rögzített számú osztályba sorolható objektumokat azonosít mikroszkópos felvételeken, akkor akár a *dashboard* átalakítás nélkül használható hozzá.

Végül, de nem utolsó sorban az eredeti algoritmusomra (vagy akár implementációmra) közvetlenül építkezve más jellegű adatokat is kinyerhetünk a mikroszkópos felvételekből, például a koordinátáinként elmentett adatokból könnyedén végezhető trajektória analízis jól ismert algoritmusok használatával.

Rövid távú tervek

Első lépésként szeretném ötvözni a *GlueFlux jelzőrendszer*hez kidolgozott *szakértői modellt* a gyűrűk keresésére irányuló próbálkozásaimmal, *mesterséges intelligenciát* alkalmazva. Jelenleg már van rá működő implementációm, a legnagyobb kihívás a megoldás általánosságának igazolása. Azt követően pedig szeretném elkészíteni ennek a dolgozatnak a biológusoknak szóló változatát és azt hazai és/vagy nemzetközi folyóiratban publikálni.

5. fejezet

Összefoglalás

A dolgozatomban részletesebben bemutatam az egyik, *szekrétációs granulumokkal* kapcsolatos kutatási eljárást, a *GlueFlux jelzőrendszer* használatát (1.3 alfejezet) és ejtettem pár szót a LampRing rendszerről is (3.5 ábra). Ezen kívül a dolgozatomban tartalmaz egy hasznos áttekintést (A. függelék) a morfológiai képelemzéssel kapcsolatos kutatások megalapozásául.

A kutatásom fő kérdése az volt, hogy pontosan hány darab *pozitív* és *negatív szekrétációs granulum* látható egy adott mikroszkópos felvételen. Bemutattam, hogy a téma még nem rendelkezik kellő mennyiségű szakirodalommal, annak ellenére, hogy aktuális (1.5. alfejezet).

Ezen kívül bemutatam egy új, a *GlueFlux jelzőrendszer* felvételeinek feldolgozására alkalmas *szakértői modellt* (2.2 alfejezet) és annak implementálására használt, gondosan összeállított szoftveres környezetet (2.1 alfejezet), aminek működését igazoltam egyrészt a sorok között részleges indoklással (leszámítva a granulumok beazonosítását), másrészt statisztikai módszerekkel (3. fejezet), mind kézi utánaszámolással, mind konkrét gyakorlati alkalmazás során.

A lehetséges további kutatási irányokat felvázoltam a sorok között és össze is foglaltam és külön kiemeltem a személyes terveimet (4. fejezet). A sikertelen kutatási kísérleteim mind egy másik rendszer, a LampRing felvételeire irányultak. Ezeket a statisztika után részleteztem (3.2 alfejezet).

5.1. Egyéni hozzájárulás

A dolgozatom az informatika és a biológia egy eddig informatikai oldalról nem vizsgált specifikus határterületével foglalkozik. A kutatás informatikai oldalát teljes mértékben egyedül, témavezetői segítség nélkül végeztem; a témavezetőm — biológusként — elsődlegesen a feladat specifikálásában segített.

Újrafeltárom a *szekrétációs granulumokkal* kapcsolatos képfeldolgozási lehetőségeket, és bemutatok egy új *szakértői modellt* a *GlueFlux jelzőrendszer* használatával készült felvételek automatizált feldolgozására. A modellem a jelenleg ismert módszerekkel ellentétben újszerű elgondolást is figyelembe vesz: Ötvözőm az objektum-alapú kolokalizációt és a hagyományos kolokalizációt azáltal, hogy diszkretizáció előtt is keresek összefüggéseket egy felvétel több csatornája között.

A modell validációja mellett azt egy valós kísérletben is alkalmazom¹: részlegesen megismétlem Roederer, Bowser és Murphy kísérletét [40] a lizoszomális enzimek által okozott savasodás hőmérséklet-függésére vonatkozóan.

¹A kísérlet végrehajtását és az adatgyűjtést Dósa Anna végezte, azonban a kvantifikálás és az adatvizualizáció az én modellemmel történt, illetve az itt leírt statisztikai elemzést önállóan végeztem

Köszönetnyilvánítás

Szeretném megköszönni a dolgozat elkészüléséhez való hozzájárulását témavezetőmnek, dr. Csizmadia Tamásnak, aki a feladat részletes, jól illusztrált specifikálása után hasznos tanácsokkal látott el az első tudományos munkám megszületésére vonatkozóan, Dósa Annának a dolgozatban és a kutatás során használt fluoreszcens mikroszkópos felvételekért, Farkas Erikának a fénymikroszkópos felvételekért. Továbbá szeretnék köszönetet mondani a Bolyai Szakszeminárium vezetőjének, Porkoláb Zoltánnak, aki akár vasárnap este is válaszolt a levelemre, és az *ELTE Biológiai intézet Anatómiai Sejt- és Fejlődésbiológiai Tanszéknek (ELTE ASF)*, azon belül is főként a CsizmadiaLab labornak [5], aminek tagjai szabadidejükben rengeteget segítettek a dolgozat biológiai háttérében amellet, hogy végig támogattak a dolgozat elkészítésében, és akiket kérdéseimmel folyton zaklathattam. Hozzájárultak a munkám színvonalához a Biológus TDK-n hallott értékes előadások és a Bolyai szakszemináriumok illetve az *ELTE ASF* tanszék inspiráló közege. Szeretném megköszönni a segítségét mindenkinek, aki félkész állapotban segített lektorálni a dolgozatomat: a családomnak, Tatai Áronnak, Dósa Annának, Kozák Balázsnak, Csizmadia Tamásnak és Hajder Leventének, akit a szakdolgozatommal kapcsolatos belső konzulensi közreműködése mellett a TDK-mmal kapcsolatos lektorálásra is megkérhettem. Továbbá mindenképpen szeretnék köszönetet mondani Csetverikov Dimitrijnek, aki a rövid találkozóink ellenére is hosszan tartó tanácsokkal látott el, és akinek közvetlenül köszönhetem, hogy a *DoG* alapú algoritmust használatba vettem.

A. függelék

Elméleti háttér

Alább részletezem azokat a fontosabb eljárásokat, algoritmusokat, amelyeket felhasználok a modellemben.

A.1. Aritmetikai műveletek képeken

Az egyik legegyszerűbb képfeldolgozási eljárás az, amikor a képen pixelenként a szín(ek)kel aritmetikai műveleteket végzünk.

Általában ezeknek csak szürkeárnyaltos (egyszínű) képek esetében van értelme.

A.1.1. Műveletek kép és skalár között

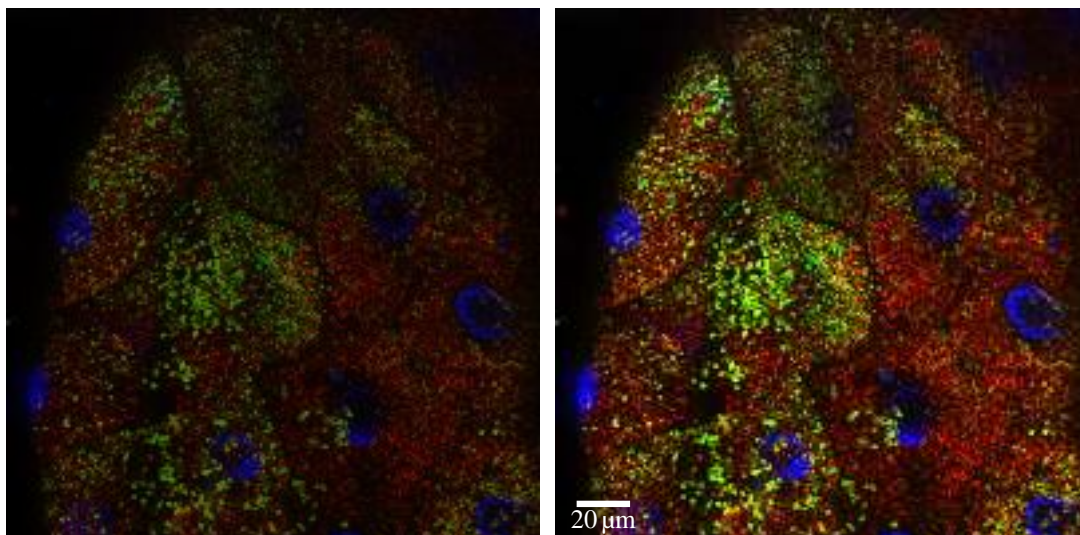
Szorzás skalárral

A legegyszerűbb ilyen művelet a skalárral való szorzás.

Ha például egy képen a legnagyobb intenzitású pixel is csak lehetséges legnagyobb intenzitás felét éri el, információvesztés nélkül felskálázható. (A.1 ábra)

Lebegőpontos formátum

Annak érdekében, hogy az információvesztést elkerüljem, olyan képformátumot használok, ami a pixelek értékét előjeles lebegőpontos számként tárolja. Mivel a lebegőpontos számábrázolás esetén a pontosság az elméleti maximumhoz közeledve csökken, a legpraktikusabb maximális intenzitásérték az 1, így könnyedén tudok az elméleti maximumnál kisebb és nagyobb fényességet is tárolni.



A.1. ábra. Kép szorzása skalárral. A jobb oldali képen színenként az intenzitás kétszer akkora, mint a bal oldalon.

Összeadás skalárral (eltolás)

Amennyiben szorzás helyett összeadást vagy kivonást alkalmazunk, az egymás melletti pixelek intenzitáskülönbsége megmarad.

Hatványozás

Ha a kép pixeleinek intenzitását négyzetre emeljük, az egymás melletti pixelek intenzitásának az aránya is változik, továbbá a nagy intenzitású pixelek esetében nagyobb az intenzitáskülönbség mint a kis intenzitású pixelek környezetében. Amennyiben az intenzitások a $[0, 1]$ tartományból származnak, ez fordítva van.

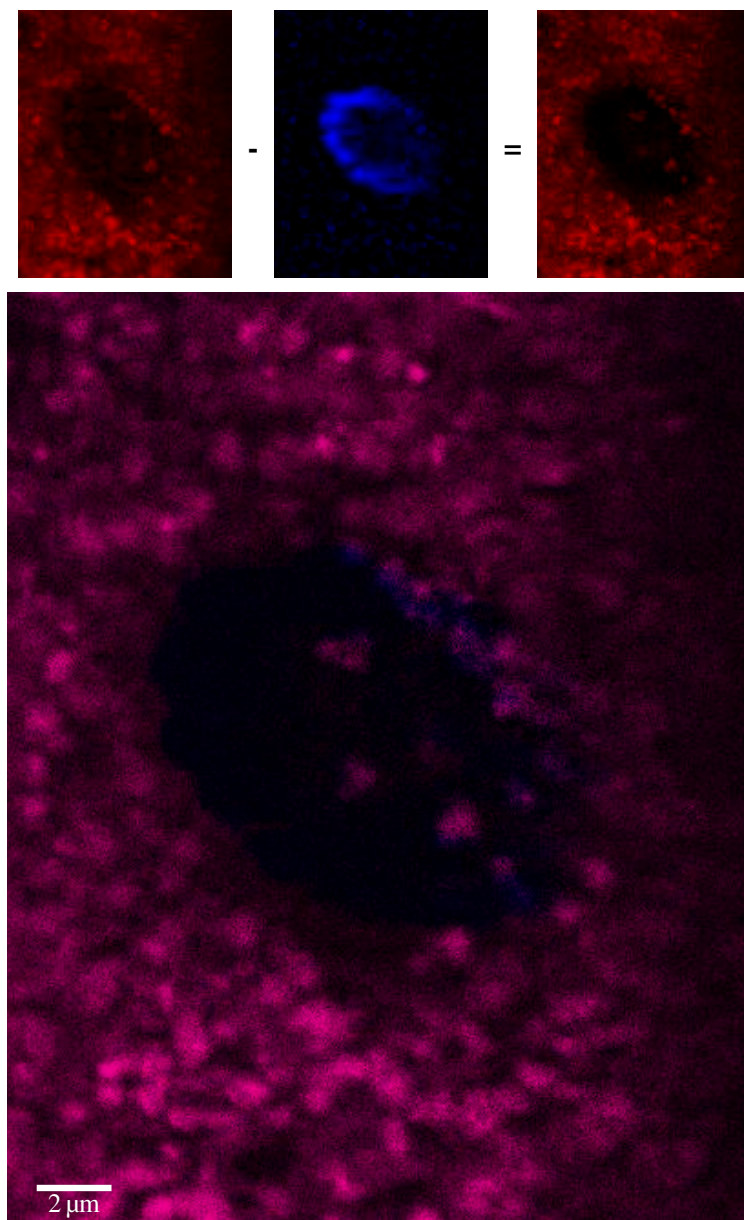
Összefoglaló az egyszerű műveletekről

Az intenzitáskülönbségek megtartása/elvesztése bizonyos esetekben kívánt tulajdonság, bizonyos esetekben pedig kerülendő.

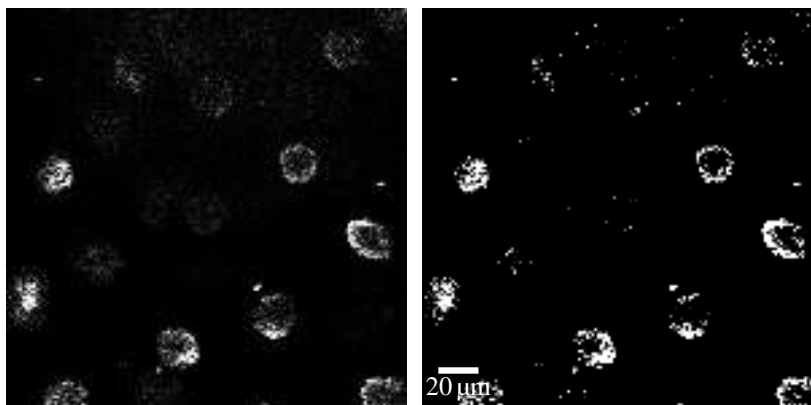
A.1.2. Műveletek képek között

Bináris aritmetikai műveletet nem csak egy adott kép pixeleivel és egy konstanssal lehet végezni, hanem két kép megfelelően párosított pixelei között is.

A képek közötti kivonás alkalmas a képek egyes részleteinek csökkentésére. [45] Például, ha egy adott képből kivonjuk ugyanarról a területről készített DAPI-szűrős változatát, eltűnnek a sejtmagok közelében lévő részletek, mivel a DAPI kizárólag a



A.2. ábra. Képek kivonása egymásból. A fenti három képen rendre ugyanannak a felvételnek a vörös (DsRed) és kék (DAPI) csatornája, illetve a kettő különbsége látható, alatta pedig a vörös csatornán a különbség, a kéken pedig az eredeti vörös csatorna. A sejtmaggal átfedésben lévő *szekekrációs granulumok* némileg kékesebbek, halványabbak.
A képek szélessége nagyjából 25μm



A.3. ábra. DAPI *fluoreszcens marker* intenzitása egy képen, limitálás előtt és után. A kör alakú formákból sejtmagok jelenlétére következtethetünk. A különbség nem szembetűnő, de ha keresünk a bal oldali képen nagyon halvány részeket, azok a jobb oldali képen hiányoznak, továbbá a világosabb részek a jobb oldalon sokkal világosabbak. A későbbiekben be fogok mutatni szembetűnőbb különbséget eredményező módszereket is.

sejtmagokban látszódik. (A.2 ábra) Azonban sajnos ez az eljárás még nem tökéletes, az ábrán az látható, hogy csak a sejtmag szélén lévő granulumok tűnnek el, és azok is csak kevésbé.

A.2. Limitálás (thresholding)

Sokkal jobb eredményt érhetünk el, ha nem-lineáris transzformációkat is alkalmazunk. Ilyen például a limitálás, amely az előjelfüggvényhez hasonlóan egy bizonyos limitáló érték alatti értékekhez 0-t, felette viszont a lehetséges legnagyobb értéket rendeli.

A A.3 ábrán az előző felvételek kék (DAPI) csatornáján egy ilyen átalakítás látható.

A 1.5 *Kapcsolódó munkák* alfejezetben említett munkák nagyrészt csak egy egyszerű limitálást végeznek a felismerendő objektumokon.

A.2.1. Otsu-féle hisztogram alapú threshold

A megfelelő limitáló érték kiválasztásához több algoritmus is létezik. Ezek közül talán a legegyszerűbb módszer, ha a limitáló értéknek a maximális és a minimális intenzitás számtani közepét vesszük, azonban ez az esetek kis részében vezet megfelelő eredményre. Az ennél jobbak közül az egyik legismertebb módszer Otsu nevéhez fűződik. Ennek lényege a határérték alatti és feletti intenzitások minimalizálása. [46]

Habár az ismertsége, könnyű implementálhatósága és átláthatósága miatt gyakran alkalmazzák például az Imaris [27] szoftverben, ez az eljárás nem feltétlenül a legjobb eredményeket hozza, helyette érdemes inkább a következő részben leírt lokális adaptív limitálást alkalmazni.

A.2.2. Lokális adaptív limitálás

Az előzőekben ismertetett ún. globális limitálási eljárásoknál általában jobb eredményeket érhetünk el, ha lokális limitálást alkalmazunk. Ennek lényege, hogy mindegyik pixel esetében más limitet alkalmazunk, ami az adott pixel egy környezetének értékeit veszi figyelembe. Ez lehetővé teszi a változó kontrasztú képek limitálását. [45]

A.3. Konvolúció

A képfeldolgozás egyik legsokoldalúbb alapvető művelete a *konvolúció*. Ez egy bináris képtranszformációs művelet képek és adatmátrixok (ún. *kernelek*) között, amely során az eredményül kapott kép összes pixelének intenzitása a környezetének és az adatmátrix között végzett skaláris szorzatnak¹ az eredménye. [48]

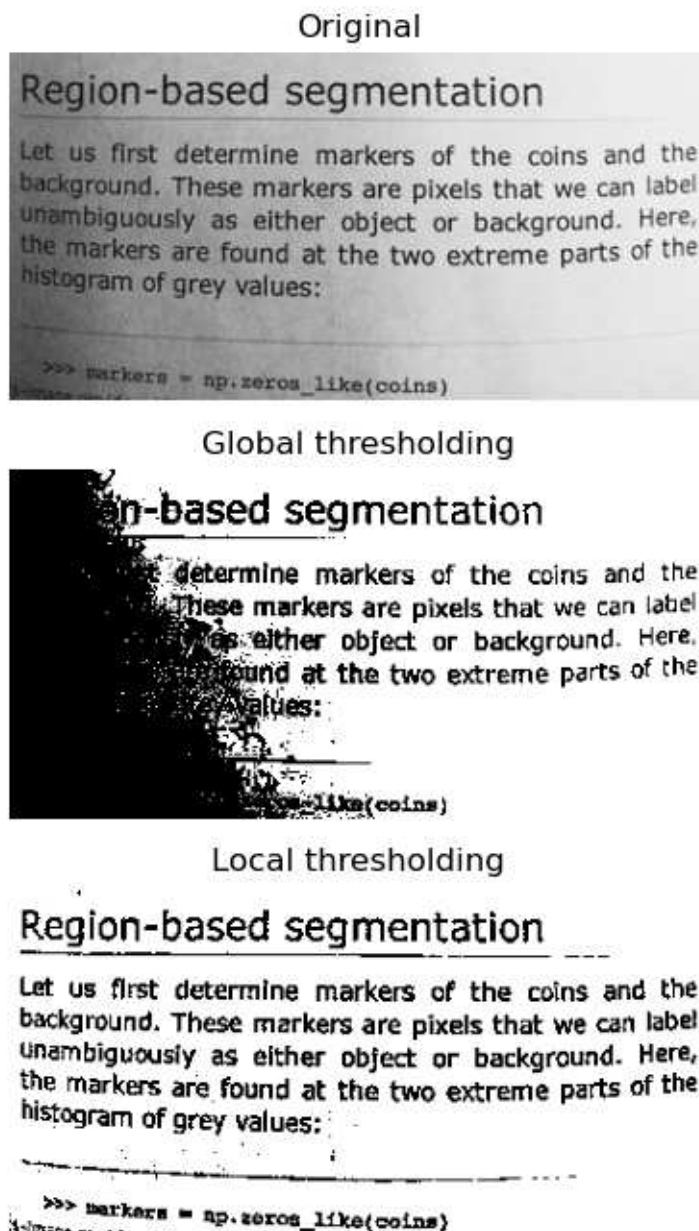
A.4. Gauss-elmosás

A *Gauss-elmosás (gaussian blur) Konvolúció* az ún. Gauss-kernellel. Az ilyen kernel vonalmenti metszetének az intenzitása a σ szórású normális eloszlás eloszlásfüggvényét követi (A.5 ábra). A kernel egy-egy eleme a A.1 képlettel írható le. [49]

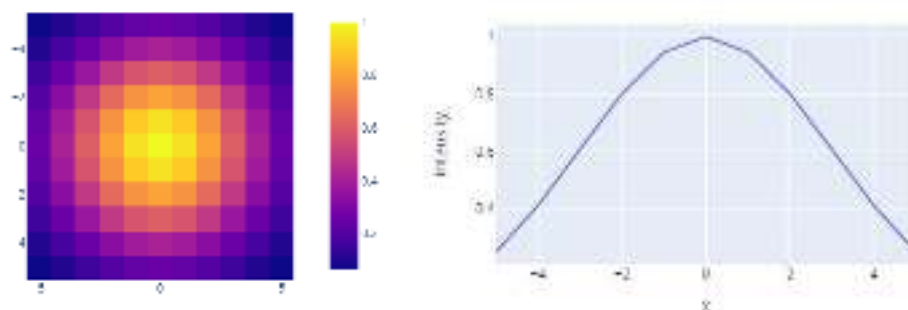
$$K[x,y] = e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (\text{A.1})$$

Ilyen kernellel az adott pixel intenzitását egyenletesen befolyásolja a környezetének intenzitása, így a kép elmosódik, a lokális elváltozások eltűnnek, csak a nagyobb területeket érintő intenzitáskülönbségek maradnak meg.

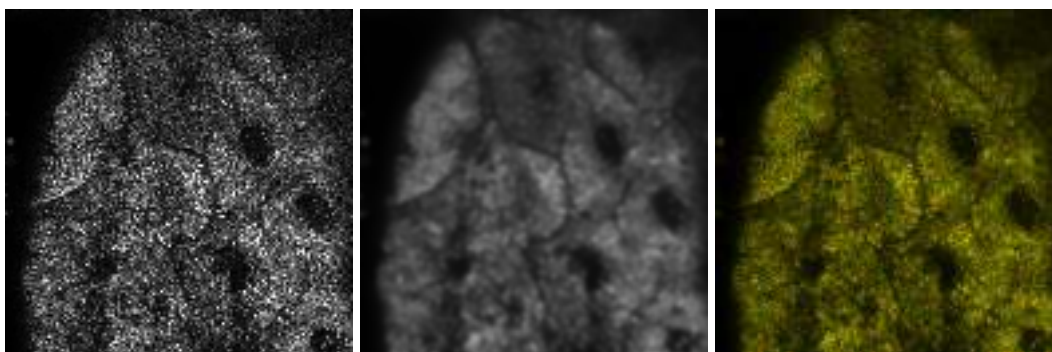
¹A skaláris szorzat elemenkénti szorzást majd a szorzatok összegzését jelenti



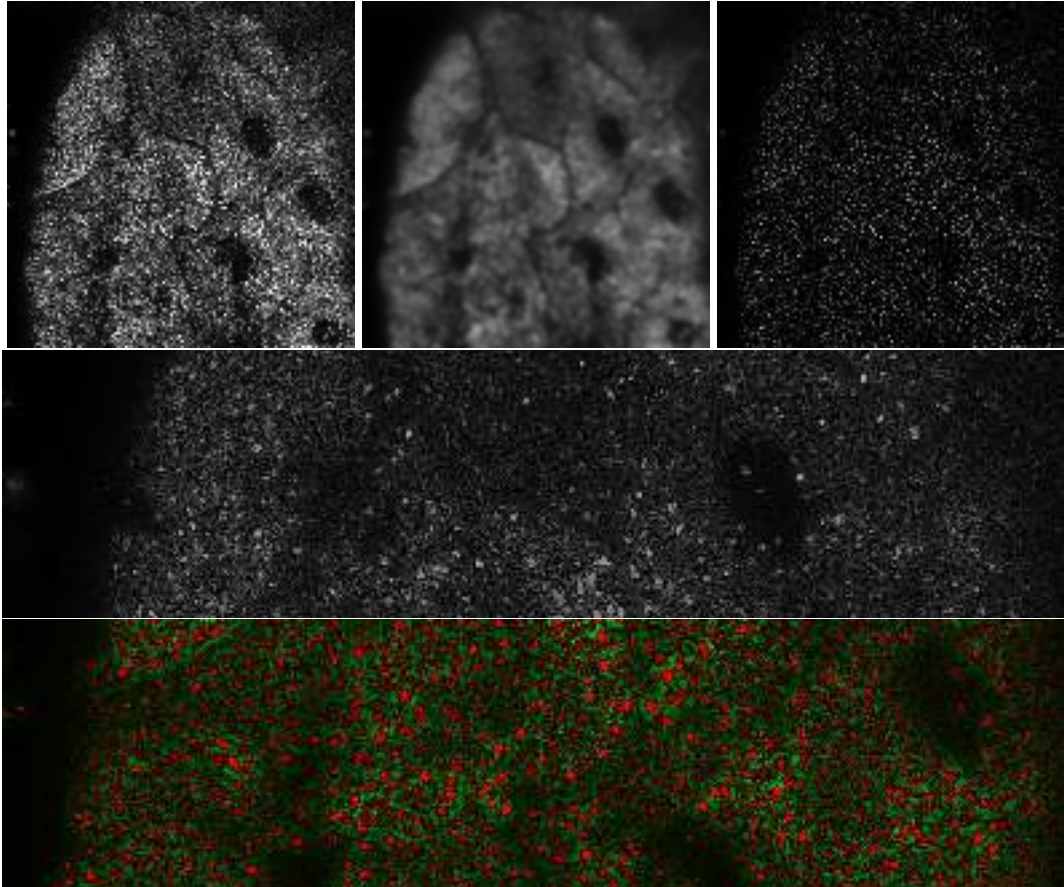
A.4. ábra. Lokális limitálás (thresholding) sokkal jobb eredményeket hoz változó kontrasztú kép esetén. A felső képen egy könyvről készült fotó látható, alatt a globális limitálás majd legalul lokális limitálás. A limitálás során a felbontás romlani látszik, de ez csak információvesztésre utal (dimenziócsökkenésre). Lokális limitálás esetén a legkevesebb a hiba. A kép forrása: skimage dokumentáció [47]



A.5. ábra. Egy 11x11-es méretű, $\sigma = 3$ paraméterű *Gauss-kernel* és annak $y = 0$ vonal menti metszete



A.6. ábra. Egy kép és annak *Gauss-elmosás* (*gaussian blur*) utáni változata. A harmadik képen a vörös csatornán az eredeti, a zöld csatornán pedig az elmosott változat látható, a képet jobban megfigyelve feltűnhet, hogy a granulumok nagy része sárga (piros + zöld), csak az átlagosnál nagyobb intenzitású granulumok kapnak pirosas árnyalatot.



A.7. ábra. A *Difference of Gaussians (DoG)* művelet szemléltetése. Egy kép eredeti, elmosott változata majd az eredeti és az elmosott különbsége. Alatta a különbség abszolút értéke látható, amelyen minden lokális elváltozás kirajzolódik. Legalul a vörös csatornán az eredetiből vonom ki az elmosottat, míg a zöld csatornán fordítva, így a vörös csatornán a lokális intenzitásnövekedések, míg a zöld csatornán a lokális intenzitáscsökkenések rajzolódnak ki.

A.4.1. *Difference of Gaussians (DoG)*

Két kép különbsége (A.1.2 rész) alkalmas lokális vonások kiemelésére vagy eltompítására. Egy elmosott képet kivonva az eredeti képből a részletek előtűnnek. (A.7 ábra)

DoG piramis

Az algoritmusom egyik központi eleme a *sarokdetektálás (blob detection)*, amely során egy 2D-s képen gömbszerű struktúrákat azonosítok. Én David G. Lowe 2004-ben publikált algoritmusát alkalmazom. Ennek legnagyobb előnye a könnyű implementálhatóság és a méret- és körvonalinvariáns detektálás. [25]

A.5. Egyéb folt-detektálási algoritmusok

Az orvosbiológia területén a leggyakrabban a *DoG*, a *LoG* és a *DoH* algoritmusokat használják, így — habár ezek a modellemben nem szerepelnek — ez utóbbi kettőről is ejtek pár szót. [50]

A.5.1. *Laplace of Gaussians (LoG)*

„A *LoG* az egyik első és leggyakrabban használt folt-detektáló [algoritmus]. A folt jellemzőt csak fehér színű foltokon képes érzékelni, és mitöbb, nagyobb foltok detektálása kifejezetten lassú. Mindegyik közül a *LoG* a legpontosabb és leglassabb eljárás.” [50]

Szerencsére a fekete/fehér színeket egy egyszerű invertálással meg lehet cserélni, azonban tapasztalatom szerint nincs szükség ennek az alkalmazására, a gyorsabb módszerrel is megfelelően pontos eredmények érhetőek el.

A.5.2. *Difference of Hessians (DoH)*

„A *LoG*-hoz hasonlóan a *DoH* a *scale-space*t veszi figyelembe a cél-pontok detektálásához. A *DoH* esetében a lényeges képek gyorsan detektálják a cél-pontokat. A *DoH* előnye a számítási igény és így a felismerő az egyik leghasznosabb *folt-detektáló*. Ha a *scale-space* nem kifejezetten nagy, a *DoH* akár valós idejű *folt-detektáláshoz* is használható. Ezentúl a detektálási sebesség független a foltok méretétől és több szokványos geometriai struktúrát is tud detektálni, például tál formákat, foltokat és lapokat. Használható mind sötét háttéren világos, mind világos háttéren sötét foltok detektálására.” [50]

Sajnos a sebesség ennek az algoritmusnak az esetében enyhén a minőség rovására megy. Tapasztaltom szerint a különbség már nem elhanyagolható, azonban amennyiben a modellem alkalmazása megköveteli, a sebesség nyugodtan növelhető *DoG* helyett *DoH* alkalmazásával, mivel a hiba teljesíti a követelményeimet (1.4.1 Milyen a jó hiba?).

A.5.3. Mintafelismerés

Egy másik, kézenfekvő módszer *folt-detektálásra* speciális algoritmusok alkalmazása helyett mintafelismerés alkalmazása. Ennek kézenfekvő módja különböző méretű granulum alakú mintával való konvolúció alkalmazása. Ez a megközelítés (a minta

személyre szabhatóságának köszönhetően) kicsit nagyobb pontosságot de egyben jóval nagyobb futásidőket is eredményez.

Készítettem erre is egy implementációt, az elgondolás működőképes, de véleményem szerint lassúsága miatt nem érdemes alkalmazni, esetleg jövőbeni kutatások esetén kerülhet szóba alkalmazása.

A.6. Morfológiai módszerek

„A matematikai morfológia olyan kép-transzformációknak és kép-függvényeknek az elmélete, amelyek halmazelméleti, geometriai és topológiai koncepciókra vannak alapozva.” [51] Ennek a megközelítésnek az előnye az, hogy a feldolgozni kívánt képek tulajdonságainak ismeretében akár vakon is kidolgozhatunk jól működő eljárásokat, morfológiai építőkövek alkalmazása esetén meglehetősen egyszerű és szemléletes az algoritmus helyes működésének indoklása, ami kifejezetten előnyös interdiszciplináris területen [26].

A morfológiában leggyakrabban használt műveletek binárisak, melyek egyik operandusa egy szürkeárnyaltos kép, a másik pedig egy bizonyos szabályszerű *kernel* összes paramétere. A valódi művelet implementálható a *kernel* és a kép összes pixele között egy elemenkénti szorzattal, majd egy halmaz-redukciós művelet (pl. minimum, maximum, összeg) alkalmazásával. [52]

A Wolfram Mathematica [53] nyelv `ListConvolve` függvénye egy általánosított *konvolúciót* valósít meg, így ideális annak a jelöléseit alkalmazni. A A.2 képlet megfelel a hagyományos konvolúciónak.

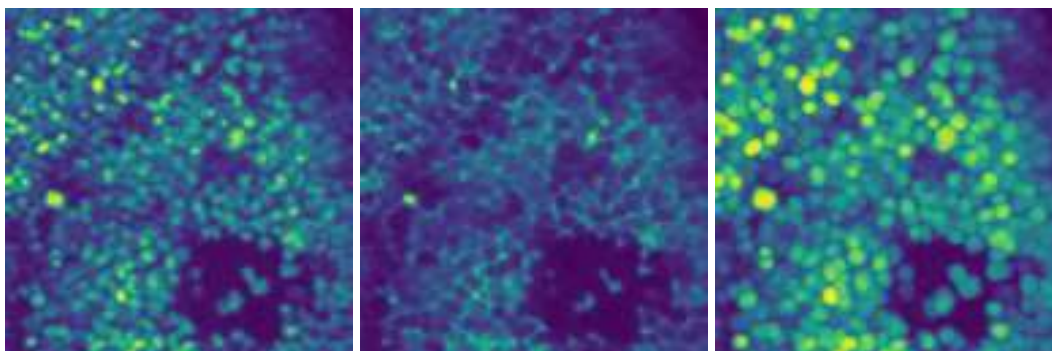
$$\text{ListConvolve}[\text{kernel}, \text{image}, \text{Ceiling}[\text{Length}[\text{kernel}]/2], \{\}, \text{Times}, \text{Plus}] // \text{MatrixForm} \quad (\text{A.2})$$

Alább ismeretek pár fontosabb morfológiai műveletet.

A.6.1. Erózió, diláció

Diláció

A *diláció* a minkowski-összeg morfológiai megvalósítása. Informálisan megfogalmazva egy képre *dilációt* alkalmazva az objektumok mérete megnő.



A.8. ábra. Az *erózió* és a *diláció*. A második ábrán az első ábra erodált változata, míg a harmadik ábrán a dilatált változata látható, 5px sugarú kör alakú kernellel.

Minden pixelnek az adott *kernelre* vonatkozó környezetéből megkeressük a maximumot. Wolfram Mathematica [53] nyelvben a A.3 képlet szerint írható le.

$$\text{ListConvolve}[\text{kernel}, \text{image}, \text{Ceiling}[\text{Length}[\text{kernel}]/2], \{\}, \text{Times}, \text{Max}] // \text{MatrixForm} \quad (\text{A.3})$$

Erózió

Az *erózió* a minkowski-különbség morfológiai megvalósítása. Informálisan megfogalmazva egy képre *eróziót* alkalmazva az objektumok mérete lecsökken.

Minden pixelnek az adott *kernelre* vonatkozó környezetéből megkeressük a minimumot. Wolfram Mathematica [53] nyelvben a A.4 képlet szerint írható le.

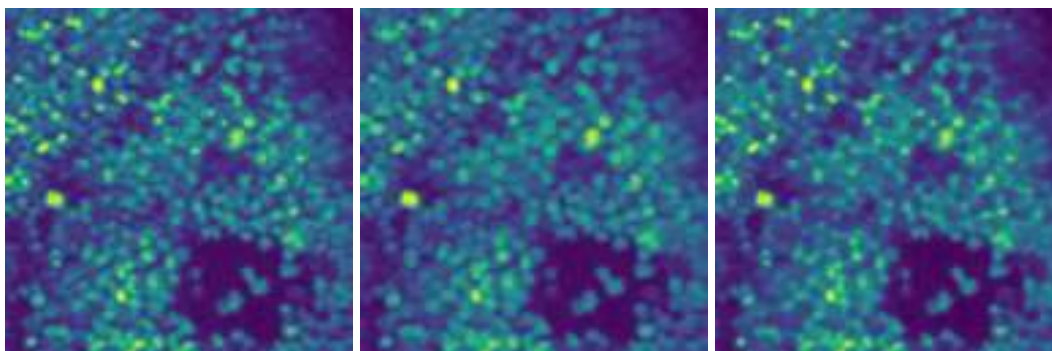
$$\text{ListConvolve}[\text{kernel}, \text{image}, \text{Ceiling}[\text{Length}[\text{kernel}]/2], \{\}, \text{Times}, \text{Min}] // \text{MatrixForm} \quad (\text{A.4})$$

A.6.2. Opening, closing

Eróziót és *dilációt* egymás után alkalmazva a kis méretű objektumok eltűnnek, ezt opening-nek hívjuk. (A.5 képlet)

Fordított sorrendben alkalmazva a kis méretű lyukak tűnnek el. (A.6 képlet)

$$\text{ListConvolve}[\text{kernel}, \text{ListConvolve}[\text{kernel}, \text{image}, \text{Ceiling}[\text{Length}[\text{kernel}]/2], \{\}, \text{Times}, \text{Min}], \text{Ceiling}[\text{Length}[\text{kernel}]/2], \{\}, \text{Times}, \text{Max}] // \text{MatrixForm} \quad (\text{A.5})$$



A.9. ábra. Az opening és a closing. A második ábrán az első ábra opening utáni változata, míg a harmadik ábrán a closing utáni változata látható, 5px sugarú kör alakú kernellel.

$$\text{ListConvolve}[\text{kernel}, \text{ListConvolve}[\text{kernel}, \text{image}, \text{Ceiling}[\text{Length}[\text{kernel}]/2], \{\}, \text{Times}, \text{Max}], \text{Ceiling}[\text{Length}[\text{kernel}]/2], \{\}, \text{Times}, \text{Min}]/\text{MatrixForm} \quad (\text{A.6})$$

A módszer működésének informális indoklása:

- opening esetében:
 1. erózió: Az objektumok mérete lecsökken, olyannyira, hogy a nagyon kicsi objektumok nyomtalanul eltűnnek
 2. diláció: Az objektumok visszanyerik eredeti méretüket
- closing esetében:
 1. diláció: Az objektumok mérete megnő, olyannyira, a közel lévő objektumok közötti űr eltűnik
 2. erózió: Az objektumok visszanyerik eredeti méretüket

Ahogy a A.8 és a A.9 ábrán látható, a módszer hátránya, hogy a kép részletessége (információtartalma) minden egyes művelet során csökken. Így a morfológiai eljárások alkalmasak alacsony entrópiájú statisztikák készítésére, de óvatosan kell kezelni az érzékeny statisztikákat.

B. függelék

Felhasználói dokumentáció

A sok tesztelési kör miatt az implementáció felhasználói felülete a szokásosnál komolyabb.

A fentiekén kívül az implementáció kifejlesztése mellett elkészült egy olyan felhasználói felület is, amivel elkészíthetők közvetlenül azok a grafikonok, amelyek egy releváns biológiai kutatás során belekerülnének akár a végső publikációba (B.3.2 dashboard)

B.1. Szoftverkövetelmények, letöltés

B.1.1. Futtatás Windows alatt

A program Windows operációs rendszer alatt futtatható változata egy egyszerűen használható önkicsomagoló program formájában kerül terjesztésre.

A bináris változat letölthető itt: <https://github.com/lezsakdomi/gfp-ring-detector/releases>
(Forráskód: github.com/lezsakdomi/gfp-ring-detector)

A legutóbbi automatikus build letölthető a következő linkkel, amennyiben az elmúlt 90 napban volt frissítés: <https://bit.ly/gfp-ring-detector-downloader>

B.1.2. Futtatás Linux alatt (csak fejlesztőknek javasolt)

A forráskódból való futtatáshoz szükség van az Anaconda telepítésére illetve egy új környezet létrehozására az `environment.yml` alapján.

Ezután az adott környezet aktiválását követően indítsuk el választásunk szerint a `server.py` vagy `interactive-server.py` programot.

B.2. Fájlformátumok

A program az AxioImager M1 és M2 programokkal képes együttműködni. Az alábbiakban a felvétel szó olyan mappákat takar, amelyeket az AxioImager M1 vagy M2 program segítségével lehet létrehozni felvételek TIFF fájlként való elmentése funkció használatával. Ilyen mappaformátum az AxioVision program 4-es verziójával is produkálható, amely könnyebben beszerezhető mint az AxioImager M1 ill. M2.

A más módon létrehozott, hasonló szerkezetű mappák hivatalosan nem támogatottak, azonban jó eséllyel működnek, egyedül a mappák és a TIFF fájlok nevére érdemes odafigyelni, ha valaki ezzel próbálkozna.

A fő program ún. `stats.txt` és `stats.csv` fájlokat hoz létre, amelyek más módon is előállíthatók, ha valaki csak a *dashboard* komponenst szeretné használni.

B.3. A program használata

B.3.1. Fő program

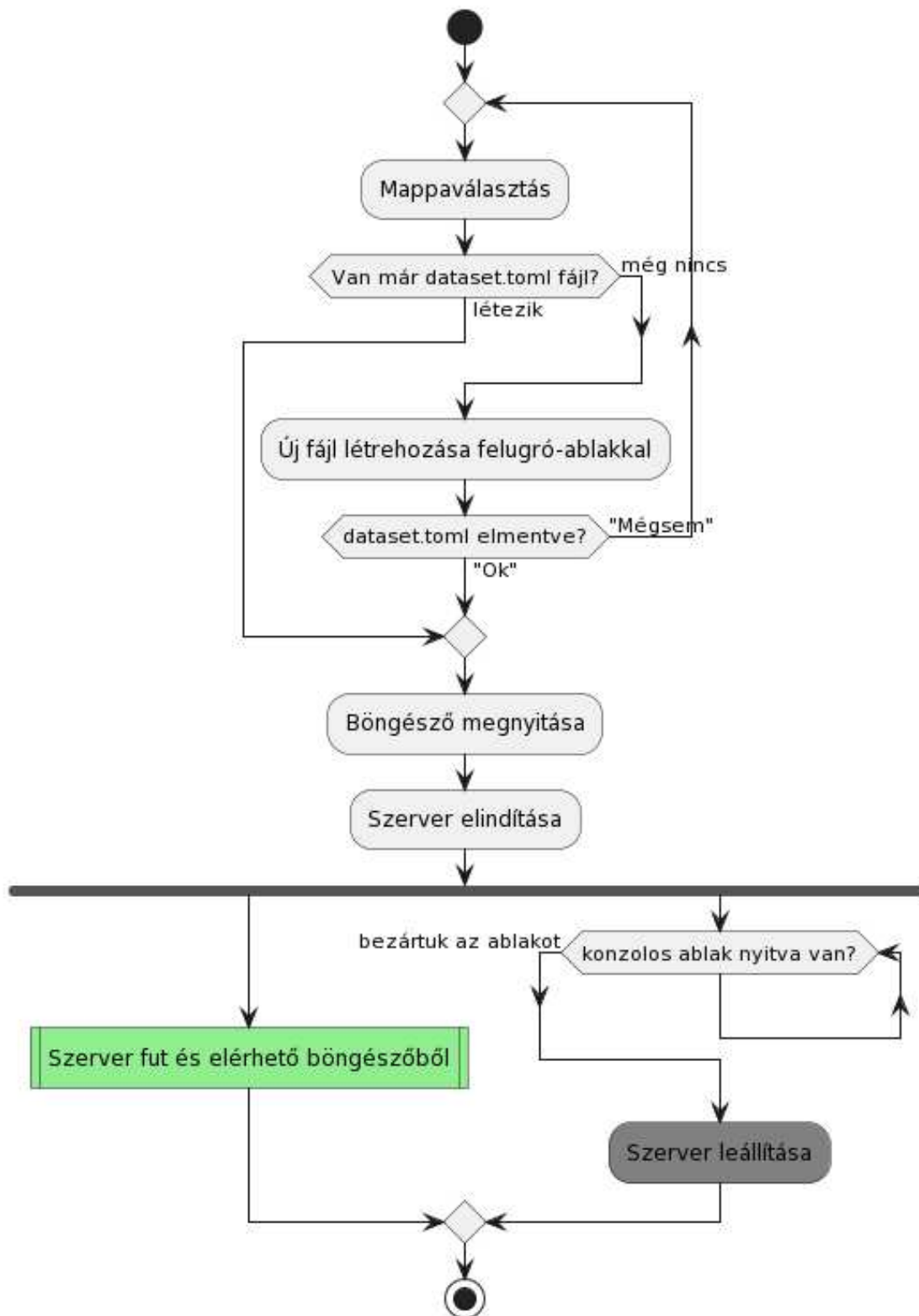
Áttekintés

A program webes felületről használható.

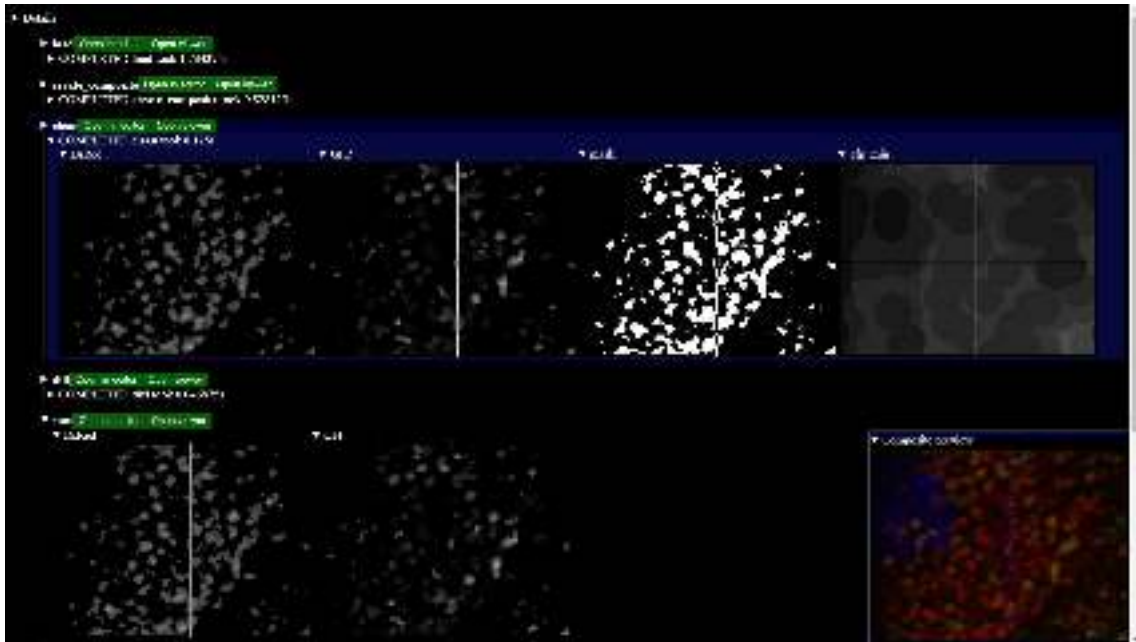
Indításkor pár felugró dialógus ablakban ki kell választanunk, hogy melyik mappában találhatóak az elemezni kívánt fájlok („dataset”), továbbá a a program megkérdez néhány, az adathalmazra vonatkozó információt, majd megnyílik egy böngészőablak, amelyben betöltődik a program webes felülete.

Ez alatt megnyílik egy konzolos ablak, amelyben a program az esetleges hibaüzeneteket élénk tárja. A program leállításához ezt az ablakot kell bezárni. Mindaddig, amíg a konzolablak nyitva van, akár több fájl is megnyithatunk különböző lapokon (B.1 ábra).

Mivel a program egy webszervert indít el, a Windows tűzfal indításkor megkérdezheti, hogy szeretnénk-e engedélyezni a más gépekkel való kommunikálást. Ha azt szeretnénk,



B.1. ábra. A letölthető program használata



B.2. ábra. Képernyőkép a fő program felhasználói felületéről

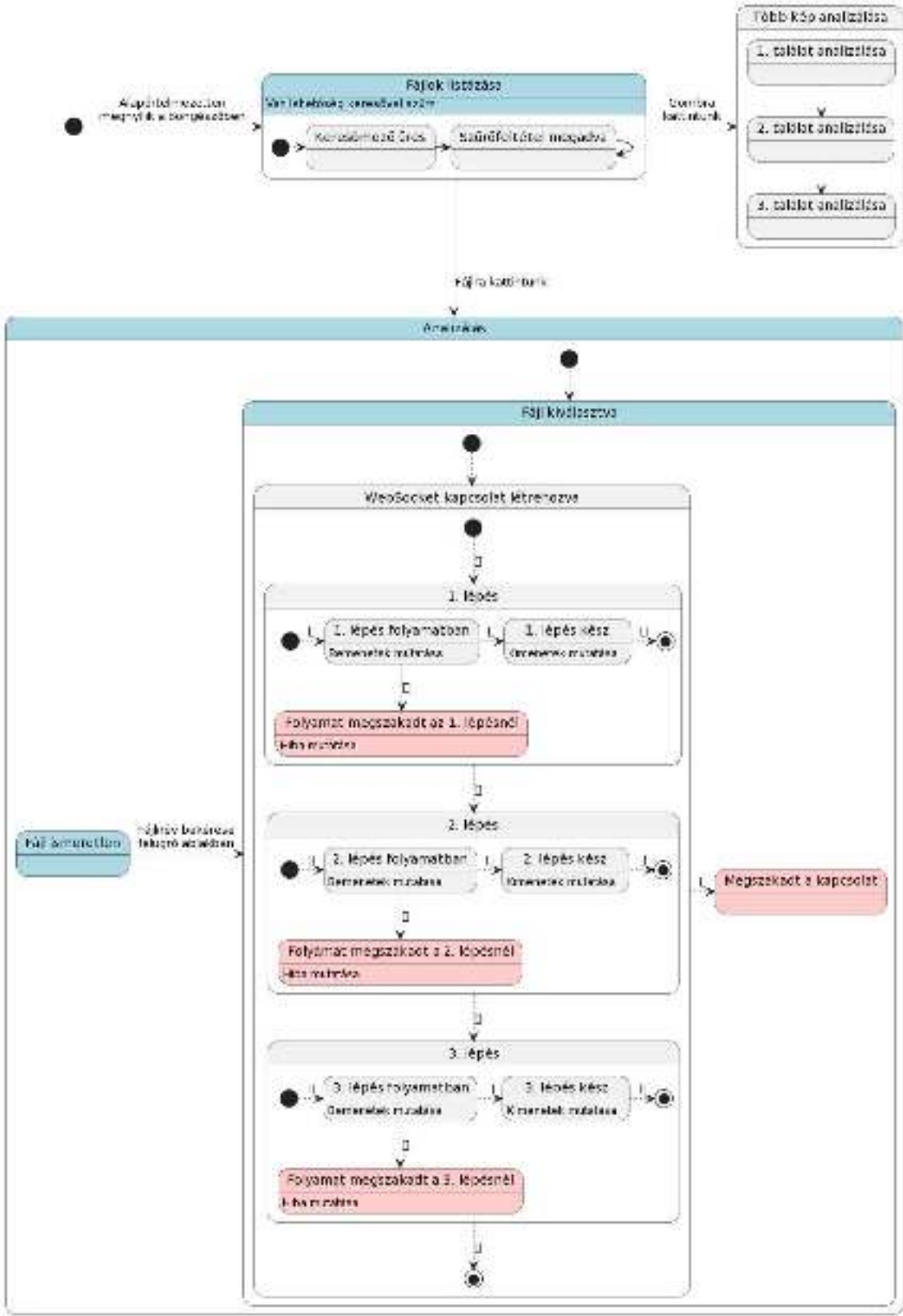
hogy a hálózatunkra kapcsolt más számítógépekről is vezérelni lehessen a programunkat, akkor engedjük át a programot a tűzfalon, ellenkező esetben ne. Sajnos a beállítás módosítása némileg körülményes, és legközelebbi indításkor nem fogja a Windows megkérdezni.

A webes felület használata

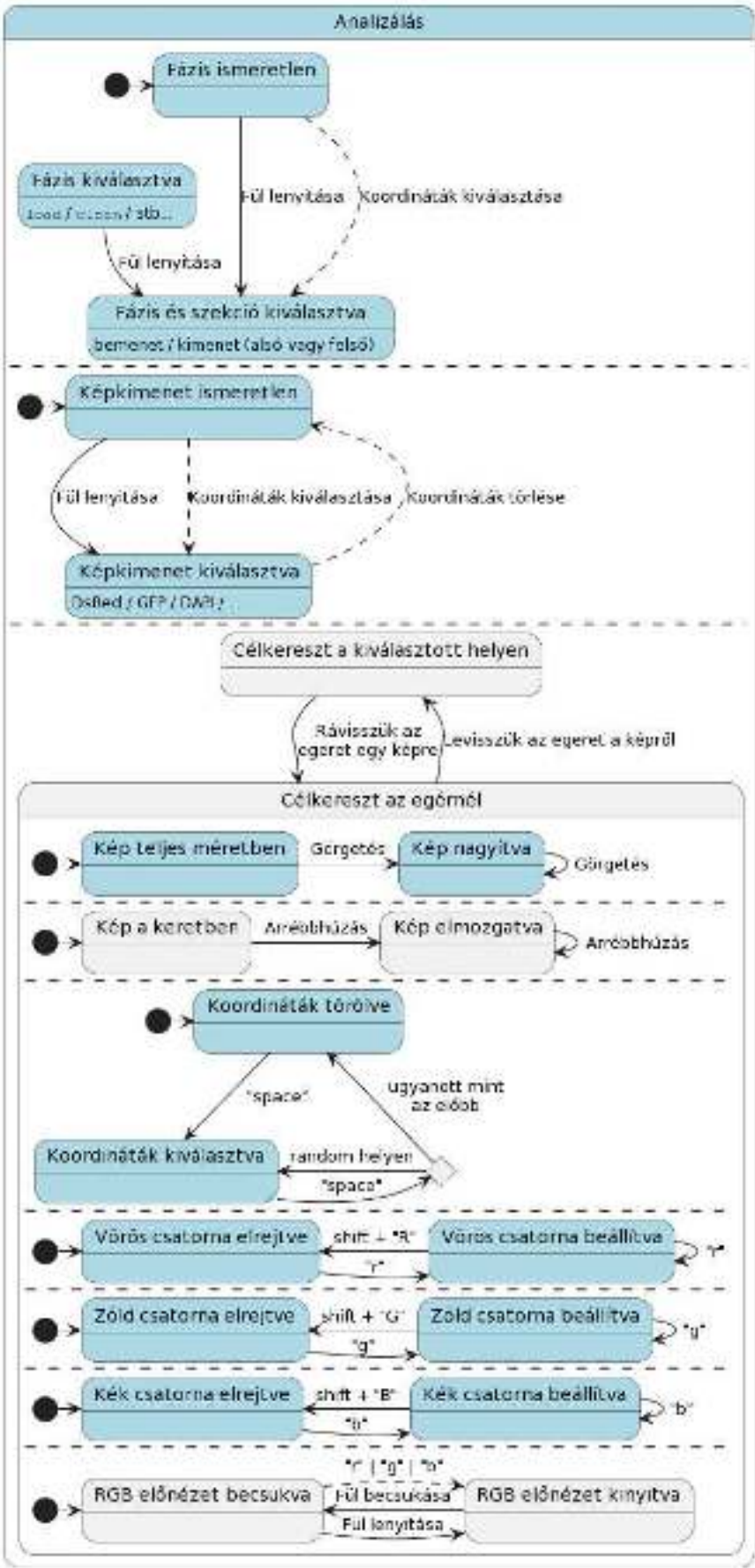
A program webes felületén lehetőség van az analizálandó felvételek listázására, egyedi képek analizálására. A fő felhasználói felület egy tipikus állapotáról készült képernyőkép látható a B.2 ábrán.

Ha kiválasztunk egy képet elemzésre, akkor egyrészt a webes felület WebSocket technológia használatával utasítja a konzolos alkalmazást analízis elkészítésére az adott képen, továbbá valós időben jeleníti meg számunkra a részeredményeket. *Mind az egyes lépések kimenetei és a bemenetei is láthatóak.*

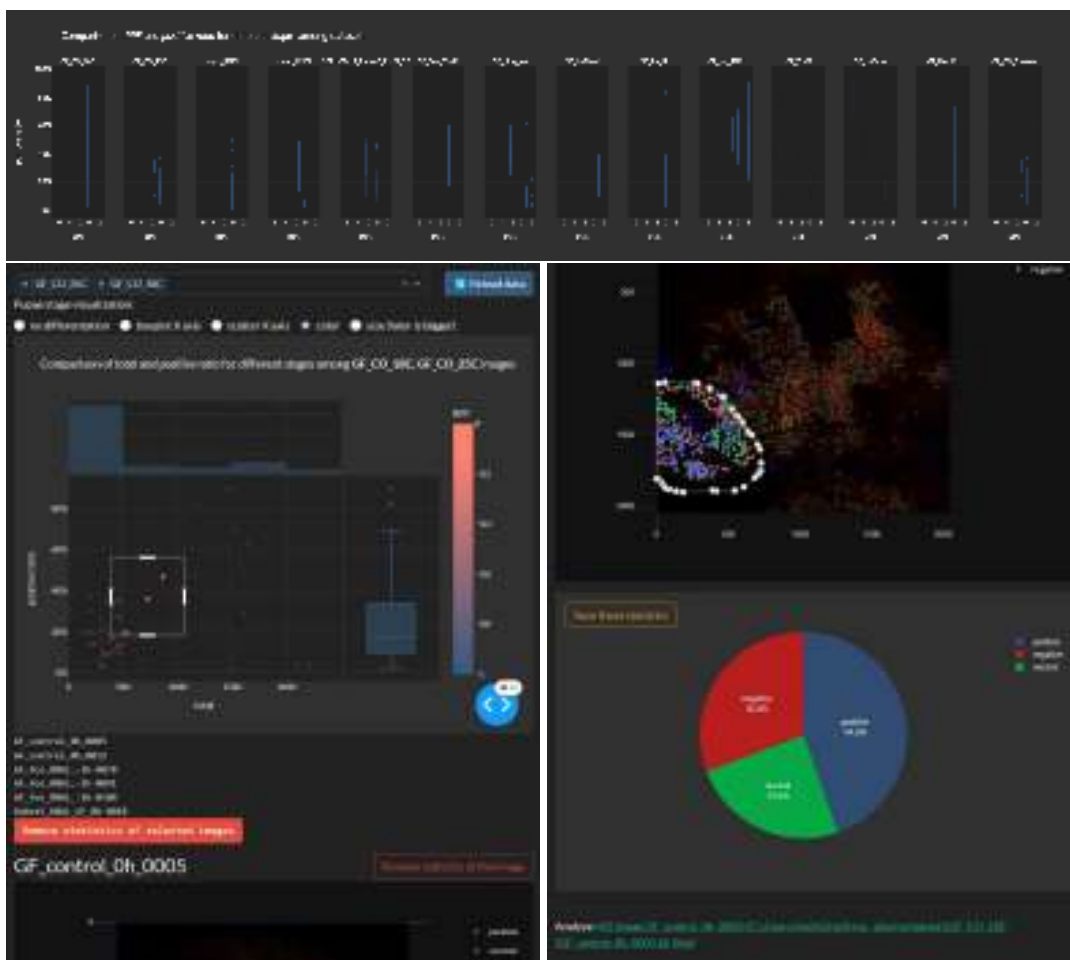
Ha rávisszük az egeret egy-egy képre, akkor van lehetőség nagyításra/kicsinyítésre (görgő), a kép arrébbhúzására, vagy akár színes kompozit létrehozására az r/g/b gombokkal, vagy a kompozit képen az egyes csatornák kikapcsolására ha ugyanezen gombok mellett nyomva tartjuk a shift gombot (B.3 ábra).



B.3. ábra. A felhasználói felület állapotai. A diagramon késsel jelöltem azokat az állapotokat, amelyek megadhatóak akkor, ha a címsort átadjuk valakinek, esetleg elmentjük az oldalt könyvjelzőbe (1/2)



B.3. ábra. A felhasználói felület állapotai. A diagramon kézzel jelöltem azokat az állapotokat, amelyek megadhatóak akkor, ha a címsort átadjuk valakinek, esetleg elmentjük az oldalt könyvjelzőbe (2/2)



B.4. ábra. Néhány képernyőkép *dashboard* komponens felhasználói felületéről

Algoritmus A program korábbi verzióiról készült egy részletesebb leírás, ez elérhető a következő linken:

<https://bit.ly/gfp-ring-detector-v2-demo>

A program által végrehajtott algoritmust az 2.2 fejezetben ismertettem (2.1 ábra).

B.3.2. dashboard

A programnak egy olyan komponense, amely segítségével grafikonon tudjuk ábrázolni és vizsgálni az eredményeket. A B.4 ábrán látható néhány képernyőmentés a programkomponens tipikus diagramjairól.

Konfiguráció

A komponens több konfigurációs paraméterrel rendelkezik. Ezeket a `config.toml` fájlban adhatjuk meg.

A **datasets kulcs** vagy egy elérési utat, vagy pedig elérési utak listáját tartalmazhatja. Lista esetén a vizsgálandó adathalmazokat jelenti, közvetlenül megadott elérési út esetén olyan mappát, amely a vizsgálandó adathalmazokat tartalmazza. Egy adathalmaz egy mappát jelent, amelynek tartalma elkülönítve jelenik meg a grafikonokon (pl. külön boxplot, külön szín, stb.).

Alapértelmezett értéke a fő program elemzésre kiválasztott mappája, ami feltételezhetően adathalmazokat tartalmaz.

Az output **kulcs** egy opcionális mappa, ahova a kimeneti fájlok (grafikon, összesített CSV fájl) kerülhetnek. Alapértelmezetten a datasets kulcs értéke kerül ide, vagy ha az egy lista, akkor a fő programnál az elemzésre kiválasztott adathalmaz.

Megnyitás

A **dashboard megnyitása a fő programból** a fájllistázós nézetben az *Open dashboard* gombra kattintva lehetséges. A kezdeti betöltés lassú, mivel a program ilyenkor átnézi az összes almappát lehetséges felvételek után kutatva és megpróbálja beolvasni az összes felvétel `stats.txt` fájlját.

A **dashboard futtatása komponensként (fejlesztőknek)** megvalósítható a `plot.py` fájlból importált `get_app()` függvény meghívásával, ami egy Plotly Dash objektumot ad vissza. Ennek a server mezője egy Flask szerver, ami rendelkezik egy `wsgi_server` mezővel, ami egyszerűen használható webes lekérések kiszolgálásához [54].

A **dashboard önálló programként való futtatása** megtehető a `plot.py` szkript megfelelő környezetben (`environment.yml` fájlban rögzített csomagok feltelepítését követően) való megnyitásával, ez azonban nem javasolt, mert így nem lehet átkattintani egy-egy pont analízisének vizsgálatára. A fő program memória- és elindulási igénye eltörpül a *dashboard* mellett.

Amennyiben a `plot.py` szkriptet önálló programként futtatjuk, az eredmények elmentésre kerülnek a konfigurációban megadott kimeneti mappában, `stats.csv` néven (amely nem követi a szokásos formátumot), és egy összefoglaló interaktív grafikon `stats.html` néven. Más módon ezen fájlok létrehozására nincs lehetőség.

Használat

A komponensben lehetőség van megadni az összehasonlítni kívánt adathalmazokat (alapértelmezetten az összes látszik), felvétel készítésének az előbábáb-állapot elérésétől számított időpont vizualizációjának módjára (ha nem választottunk ki adathalmazokat, akkor erre kevesebb lehetőségünk van a grafikon átláthatatlanságának elkerülése érdekében, ha valaki mégis azt szeretné, akkor ki kell választania az összes adathalmazt).

A megjelenített grafikonok letölthetőek képként a grafikonok jobb felső sarkában lévő kamera ikonra kattintva, amely (egyéb hasznos ikonok mellett) csak akkor látszik, ha fölé visszük az egerünket. A grafikonok interaktívak, az egeret az elemek fölé húzva részletes információkat látunk és a jelmagyarázatban kattintva egy-egy adatsort könnyedén lehet elrejteni és megmutatni. Az egyes adatpontok kattinthatóak és kijelölhetőek, kijelölés előtt a szaggatott négyzet vagy lasszó ikonra kattintva az adott grafikont kijelölés módba kell léptetni.

A felső (felvételeket összesítő) grafikonon

Adatpontok kijelölése esetén vágólapra másolható formában a grafikon alatt megjelenik a kijelölt adatpontokhoz tartalmazó felvétel neve és egy gomb, amellyel eltávolíthatjuk azokat a statisztikából. A statisztikából való eltávolítás az adott felvételekhez tartozó `stats.txt` fájl törlését jelenti. A program újraindítása vagy a jobb felső sarokban lévő *Reload data* gombra kattintásig az adatok a grafikonokon maradnak. A gomb használata közben egy felugró figyelmeztetés jelenik meg a művelet destruktív jellegéről, ami lehetőséget ad a felhasználónak arra, hogy még meggondolja magát.

Adatpontra kattintás esetén kis várakozás után (ami közben a felhasználói élmény javítása érdekében egy valós információkon alapuló folyamatjelző sáv jelenik meg ha a folyamat hosszabban tart) a grafikon alatt egy újabb grafikon jelenik meg, amin láthatjuk az adatponthoz tartozó felvétel RGB kompozit képét, amire (az adott felvételhez tartozó `stats.csv` fájl létezése esetén) felkerülnek adatpontként a program által felismert granulumok is. Továbbá az alsó grafikon jobb felső sarkában megjelenik egy gomb, amellyel a kijelöléshez hasonló módon ki lehet zárni az adott felvételt a statisztikából.

A felvételt ábrázoló grafikon sikeres betöltése esetén az alatt megjelenik egy kördiagram is, amelyen a granulumok osztályozását láthatjuk. Ennek elsősorban a

felvételt ábrázoló grafikonon adatpontok kijelölése esetén van szerepe.

Az oldal alján megjelenik egy link, amellyel a fő programban elindíthatjuk a felvétel megismételt analízisét, ezáltal a felvételhez tartozó `stats.txt` és `stats.csv` alaphelyzetbe állítását.

A középső (felvételt ábrázoló) grafikonon

Adatpontok kijelölése esetén az alatta lévő kördiagramon a `stats.csv` fájlból betöltött adatok helyett a kijelölt adatpontok eloszlása jelenik meg. A kördiagram bal felső sarkában lévő gombra kattintva felülbíráلhatjuk az adott felvételhez tartozó `stats.txt` fájl tartalmát olyanra, hogy a kördiagramon mutatott információkat reprezentálja. Azonban fontos megjegyezni, hogy a fő program és a *dashboard* független, így a gombra mentés merőben más jellegű adatokat ment el, mint amik abba automatikusan belekerülnek. A kép kizárásának esetéhez hasonlóan, a program a művelet végrehajtása közben figyelmeztet a művelet destruktív jellegére.

Grafikon területén való kattintás esetén a megfelelő koordinátákra fókuszálva megnyílik a fő programban az adott felvétel analízise, az esetleges hibák felderítésének az érdekében.

Az alsó (kördiagram) grafikon nem tartalmaz kattintható és/vagy kijelölhető adatpontokat.

Irodalomjegyzék

- [1] Attila Boda. “A *Drosophila* Arl8 kis GTPáz szerepe a lizoszomális fúziós folyamatokban”. Dissz. Eötvös Loránd Tudományegyetem, 2020. URL: https://edit.elte.hu/xmlui/bitstream/handle/10831/57032/doktori_ertekezes_BodaAttila.pdf.
- [2] Brendan Reed és tsai. “Lysosomal cathepsin creates chimeric epitopes for diabetogenic CD4 T cells via transpeptidation”. *Journal of Experimental Medicine* 218.2 (2020). DOI: 10.1084/jem.20192135.
- [3] Kevin Vivot és tsai. “Breaking Bad and Breaking Good: -Cell Autophagy Pathways in Diabetes”. *Journal of Molecular Biology* 432.5 (2020), 1494–1513. old. DOI: 10.1016/j.jmb.2019.07.030.
- [4] Walter Halangk és tsai. “Role of cathepsin B in intracellular trypsinogen activation and the onset of acute pancreatitis”. *Journal of Clinical Investigation* 106.6 (2000), 773–781. old. DOI: 10.1172/jci9411.
- [5] *Csizmadia Lab – Krinofágia kutatócsoport*. 2022. URL: <https://csizmadialab.hu/> (elérés dátuma 2022. 11. 17.).
- [6] Robert E. Smith és Marilyn G. Farquhar. “LYSOSOME FUNCTION IN THE REGULATION OF THE SECRETORY PROCESS IN CELLS OF THE ANTERIOR PITUITARY GLAND”. *Journal of Cell Biology* 31.2 (1966), 319–347. old. DOI: 10.1083/jcb.31.2.319.
- [7] G. FRAENKEL és VICTOR J. BROOKES. “THE PROCESS BY WHICH THE PUPARIA OF MANY SPECIES OF FLIES BECOME FIXED TO A SUBSTRATE”. *The Biological Bulletin* 105.3 (1953), 442–449. old. DOI: 10.2307/1538461.
- [8] Manon Monier és Virginie Courtier-Orgogozo. “*Drosophila* Glue: A Promising Model for Bioadhesion”. *Insects* 13.8 (2022), 734. old. DOI: 10.3390/insects13080734.

- [9] Anna Dósa. “A K63-kapcsolt poliubiquitin lehetséges szerepe a szekréciós granulum-lizoszóma fúzióban”. *BTDK Konferencia*. ELTE Biológus TDK, 2022.
- [10] Andrea Weckman és tsai. “Autophagy in endocrine tumors”. *Endocrine-Related Cancer* 22.4 (2015), R205–R218. DOI: 10.1530/erc-15-0042.
- [11] Tal Rouso, Eyal D. Schejter és Ben-Zion Shilo. “Orchestrated content release from *Drosophila* glue-protein vesicles by a contractile actomyosin network”. *Nature Cell Biology* 18.2 (2015), 181–190. old. DOI: 10.1038/ncb3288.
- [12] Jason Burgess és tsai. “AP-1 and clathrin are essential for secretory granule biogenesis in *Drosophila*”. *Molecular Biology of the Cell* 22.12 (2011). Szerk. Anne Spang, 2094–2105. old. DOI: 10.1091/mbc.e11-01-0054.
- [13] Tamás Csizmadia és tsai. “Molecular mechanisms of developmentally programmed crinophagy in *Drosophila*”. *Journal of Cell Biology* 217.1 (2017), 361–374. old. DOI: 10.1083/jcb.201702145.
- [14] Talley J. Lambert. “FPbase: a community-editable fluorescent protein database”. *Nature Methods* 16.4 (2019), 277–278. old. DOI: 10.1038/s41592-019-0352-8.
- [15] Arwed Weigel, Detlev Schild és Andre Zeug. “Resolution in the ApoTome and the confocal laser scanning microscope: comparison”. *Journal of Biomedical Optics* 14.1 (2009), 14022. old. DOI: 10.1117/1.3083439.
- [16] Tamás Csizmadia és tsai. “Developmental program-independent secretory granule degradation in larval salivary gland cells of *Drosophila*”. *Traffic* (2022). DOI: 10.1111/tra.12871.
- [17] *imagej count objects - Google Search*. 2022. dec. 1-én, ELTENET-ről, inkognitó módú böngészőből elvégzett keresés alapján. 2022. URL: <https://www.google.com/search?q=imagej+count+objects> (elérés dátuma 2022. 12. 01.).
- [18] *ImageJ dokumentáció - Particle Analysis*. 2022. URL: <https://imagej.net/imaging/particle-analysis> (elérés dátuma 2022. 12. 01.).
- [19] Bognár Gergő. *Modell-alapú neurális hálók*. 2022. URL: https://gsd.web.elte.hu/lectures/bolyai/2022/modell/BG_Bolyai_20221019.pdf (elérés dátuma 2022. 12. 01.).
- [20] Arató Miklós. *Matematikai Statisztika előadás*. <https://amiklos.web.elte.hu/Oktatas/Matstat2022osz/>. Hozzáférés dátuma: 2022-12-01. 2022.
- [21] Giulia De Luca és tsai. “Re-scan confocal microscopy (RCM) improves the resolution of confocal microscopy and increases the sensitivity”. *Methods and*

- Applications in Fluorescence* 5.1 (2017), 15002. old. DOI: 10.1088/2050-6120/5/1/015002.
- [22] Dmitry Chetverikov. “COMPUTER VISION: GOALS, MEANS, AND PROBLEMS”. *Computer Analysis of Images and Patterns*. Szerk. Klaus Voss, Dmitry Chetverikov és Gerald Sommer. Berlin, Boston: De Gruyter, 1989, 11–14. old. ISBN: 9783112576205. DOI: doi:10.1515/9783112576205-004.
- [23] Jesse S. Aaron, Aaron B. Taylor és Teng-Leong Chew. “Image co-localization – co-occurrence versus correlation”. *Journal of Cell Science* 131.3 (2018). DOI: 10.1242/jcs.211847.
- [24] S. Bradbury. “Microscopical Image Analysis: Problems And Approaches*”. *Journal of Microscopy* 115.2 (1979), 137–150. old. DOI: 10.1111/j.1365-2818.1979.tb00163.x.
- [25] David G. Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. *International Journal of Computer Vision* 60.2 (2004), 91–110. old. DOI: 10.1023/b:visi.0000029664.99615.94.
- [26] Mahdi Khosravy és tsai. “Morphological Filters: An Inspiration from Natural Geometrical Erosion and Dilation”. *Nature-Inspired Computing and Optimization*. Springer International Publishing, 2017, 349–379. old. DOI: 10.1007/978-3-319-50920-4_14.
- [27] Oxford Instruments. *Imaris (RRID:SCR_007370)*. URL: <https://imaris.oxinst.com/> (elérés dátuma 2022. 11. 17.).
- [28] Cheng-I Ma és Julie Brill. “Quantitation of Secretory Granule Size in Drosophila Larval Salivary Glands”. *BIO-PROTOCOL* 11.11 (2021). DOI: 10.21769/bioprotoc.4039.
- [29] Quorum Technologies. *Volocity Software*. 2021. URL: <https://www.volocity4d.com/> (elérés dátuma 2022. 11. 17.).
- [30] Shirin Nasr-Isfahani, Atefeh Mirsafian és Ali Masoudi-Nejad. “A New Approach for Touching Cells Segmentation”. *2008 International Conference on BioMedical Engineering and Informatics*. 1. köt. 2008, 816–820. old. DOI: 10.1109/BMEI.2008.262.
- [31] Caroline A Schneider, Wayne S Rasband és Kevin W Eliceiri. “NIH Image to ImageJ: 25 years of image analysis”. *Nature Methods* 9.7 (2012), 671–675. old. DOI: 10.1038/nmeth.2089.

- [32] Travis E. Oliphant. “Python for Scientific Computing”. *Computing in Science & Engineering* 9.3 (2007), 10–20. old. ISSN: 1558-366X. DOI: 10 . 1109 / MCSE . 2007 . 58.
- [33] Stefan van der Walt, S. Chris Colbert és Gael Varoquaux. “The NumPy Array: A Structure for Efficient Numerical Computation”. *Computing in Science and Engg.* 13.2 (2011), 22–30. ISSN: 1521-9615. DOI: 10 . 1109 / MCSE . 2011 . 37. URL: <https://doi.org/10.1109/MCSE.2011.37>.
- [34] Stefan Behnel és tsai. “Cython: The Best of Both Worlds”. *Computing in Science and Engg.* 13.2 (2011), 31–39. ISSN: 1521-9615. DOI: 10.1109/MCSE.2010.118.
- [35] Siu Kwan Lam, Antoine Pitrou és Stanley Seibert. “Numba: A LLVM-Based Python JIT Compiler”. *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC. LLVM '15*. Austin, Texas: Association for Computing Machinery, 2015. ISBN: 9781450340052. DOI: 10 . 1145 / 2833157 . 2833162.
- [36] Áron Kozák. “Buborékdinamika a hidrogéngyártásban: energetikai hatékonyság numerikus optimalizációja”. *Gépészmérnöki Kar, Áramlástan szekció. BME Tudományos Diákkör*, 2022. URL: <https://tdk.bme.hu/GPK/AT11/Buborekdinamika-a-hidrogengyartasban1>.
- [37] Claire McQuin és tsai. “CellProfiler 3.0: Next-generation image processing for biology”. *PLOS Biology* 16.7 (2018). Szerk. Tom Misteli, e2005970. DOI: 10 . 1371/journal.pbio.2005970.
- [38] Evan Czaplicki. “Elm: Concurrent frp for functional guis”. *Senior thesis, Harvard University* 30 (2012).
- [39] Evan Czaplicki. *An Introduction to Elm*. 2018.
- [40] Mario Roederer, Robert Bowser és Robert F. Murphy. “Kinetics and temperature dependence of exposure of endocytosed material to proteolytic enzymes and low pH: Evidence for a maturation model for the formation of lysosomes”. *Journal of Cellular Physiology* 131.2 (1987), 200–209. old. DOI: 10 . 1002 / jcp . 1041310209.
- [41] Kenneth W. Dunn, Malgorzata M. Kamocka és John H. McDonald. “A practical guide to evaluating colocalization in biological microscopy”. *American Journal of Physiology-Cell Physiology* 300.4 (2011), C723–C742. old. DOI: 10 . 1152 / ajpcell.00462.2010.

- [42] HK Yuen és tsai. “Comparative study of Hough Transform methods for circle finding”. *Image and Vision Computing* 8.1 (1990), 71–77. old. DOI: 10.1016/0262-8856(90)90059-e.
- [43] T.J. Atherton és D.J. Kerbyson. “Size invariant circle detection”. *Image and Vision Computing* 17.11 (1999), 795–803. old. ISSN: 0262-8856. DOI: 10.1016/S0262-8856(98)00160-7.
- [44] Daniel Kondermann. “Ground Truth Design Principles: An Overview”. *Proceedings of the International Workshop on Video and Image Ground Truth in Computer Vision Applications*. VIGTA '13. St. Petersburg, Russia: Association for Computing Machinery, 2013. ISBN: 9781450321693. DOI: 10.1145/2501105.2501114.
- [45] Stan Z. Li. *Encyclopedia of Biometrics*. en. Springer US, 2009. ISBN: 978-0-387-73002-8.
- [46] Nobuyuki Otsu. “A Threshold Selection Method from Gray-Level Histograms”. *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), 62–66. old. DOI: 10.1109/TSMC.1979.4310076.
- [47] *Thresholding – skimage v0.19.2 docs*. Felhasználási feltételek: <https://scikit-image.org/docs/0.19.x/license.html>. 2022. URL: https://scikit-image.org/docs/stable/auto_examples/applications/plot_thresholding.html (elérés dátuma 2022. 11. 23.).
- [48] Vincent Dumoulin és Francesco Visin. *A guide to convolution arithmetic for deep learning*. 2016. DOI: 10.48550/ARXIV.1603.07285.
- [49] Moo K. Chung. *Gaussian kernel smoothing*. 2020. DOI: 10.48550/ARXIV.2007.09539.
- [50] Kay Thwe Min Han és Bunyarit Uyyanonvara. *A Survey of Blob Detection Algorithms for Biomedical Images*. 2016. DOI: 10.1109/ictemsys.2016.7467122.
- [51] Henk J. A. M. Heijmans. “Mathematical Morphology: A Modern Approach in Image Processing Based on Algebra and Geometry”. *SIAM Review* 37.1 (1995), 1–36. old. DOI: 10.1137/1037001.
- [52] Mariusz Jankowski. “Erosion, dilation and related operators”. *8th International Mathematica Symposium*. 2006, 1–10. old. URL: http://www.internationalmathematicasymposium.org/IMS2006/IMS2006_CD/articles/Jankowski.pdf (elérés dátuma 2022. 11. 24.).

- [53] Wolfram Research, Inc. *Mathematica, Version 13.1*. Champaign, IL, 2022. URL: <https://www.wolfram.com/mathematica>.
- [54] P.J. Eby. *Python Web Server Gateway Interface*. Techn. jel. 3333. 2010. URL: <https://peps.python.org/pep-3333/>.
- [55] J.K. Ousterhout. “Scripting: higher level programming for the 21st Century”. *Computer* 31.3 (1998), 23–30. old. ISSN: 1558-0814. DOI: 10.1109/2.660187.
- [56] A. Kaspers. “Blob Detection”. Dipl. Utrecht University, 2011. DOI: 20.500.12932/6944.
- [57] Tom M. Apostol. *Calculus*. en. John Wiley & Sons, 1967. ISBN: 0471000051.

Ábrák jegyzéke

- 1.1. A kvantifikációs probléma, a különböző színű foltok számolása. Oszloponként ugyanannak a háromszínű felvételnek három különböző projekciója látható: A felső képen az összes szín látható, amit kettébontottam: A második sorban csak a magenta (*szekrétációs granulumokat* jelölő) és a kék (sejtmagokat jelölő) színt komponens, az alsó sorban azonban csak a zöld (*pozitív szekrétációs granulumokat* jelölő) színt komponens szerepel. Az ábráról szemrevételezéssel is leolvasható, hogy a bal oldalhoz képest a jobb oldalon több magenta színben látható folt zöld színben is látszik. A kutatásom fő kérdése: Pontosan hány darab folt (*szekrétációs granulum*) látható, és ezek közül mennyi zöld (kísérlet szempontjából *pozitív* minta) is és mennyi csak magenta színű (kísérlet szempontjából *negatív* minta)? (A felvételek fluoreszcens mikroszkóppal készültek kontroll (bal oldal) és géncsendesített ($Vps41^{RNAi}$ -kezelt) (jobb oldal) fiatal *Drosophila melanogaster* bábok nyálmirigy-sejtjeiről. A magenta szín DsRed, a zöld szín GFP és a kék szín DAPI jelét mutatja [5]). 5

- 1.2. A kísérleti modellállat bemutatása. Balról jobbra sorban először egy muslicák fenntartására szolgáló üvegcsövet, majd abban pár kifejlődés előtt álló lárvát, utána lárvák nyálmirigyeit láthatjuk különböző nagyításokban. A képeken lévő állatok transzgénikusak, nyálmirigyeik magenta színűnek látszanak, ami a középső képen szabad szemmel is kivehető. A jobb szélső, nagyobb nagyítású képek hasonlóak, de azokon csak a *szekrétációs granulumok* beltartalma került géntechnológiával megjelölésre, így a granulumok elkülöníthetőek, cserébe a magenta szín csak speciális mikroszkópon keresztül látható. Néhány granulum zöld színben is feltűnik egy másik, mesterséges jelölésnek köszönhetően. (A kék vonalak tájékoztató jellegűek, mindegyik képen másik egyedről készült felvétel. A fekete méretvonallal jelölt képeken +;fkh-Gal4,UAS-6XmCherry transzgénikus *Drosophila melanogaster* lárvák láthatóak (háttérben sötét színnel két fiatal, kifejlett példány), ennek köszönhetően a nyálmirigyeikben lévő granulumok magenta színben világítanak már látható fényben is. A fehér méretvonallal jelölt képeken hasonló (*GlueFlux* jelzőrendszerrel ellátott) lárvák kiboncolt nyálmirigye látható fluoreszcens mikroszkóp alatt, aminek az esetében a kibocsájtott fény valójában 575 nm-es hullámhosszon a legerősebb, a magenta színt csak a kontrasztarány növelése érdekében végzett szoftveres utószínezésnek köszönhetően kapjuk.) 7
- 1.3. A *krinofágia* folyamata leegyszerűsítve. A szekrétációs granulumok a szekrétó (normális folyamat) helyett a *krinofágia* útjára is léphetnek [9] . 8
- 1.4. Az általam megfigyelt kísérletek során leggyakrabban használt *fluoreszcens markerrendszer*. Kék, zöld és narancssárga színnel rendre a DAPI, GFP és DsRed jelölőanyagok fénykibocsájtási spektruma látható egy diagramon ábrázolva. Az én képeimen csak ez a rendszer van használatban, más kísérletek esetén más *fluoreszcens markerrendszer* is előfordulhat. A grafikon az FPbase Spectra Viewer segítségével készült [14]. (A vízszintes tengelyen a fény hullámhossza szerepel 375-től 750 nm-ig, a függőleges tengelyen pedig az egyes anyagok fluoroflórjai által az adott hullámhosszon kibocsájtott fény (normalizált) intenzitása látható) 10

1.5. <i>Fluoreszcens</i> markerek egyesítése. <i>Drosophila melanogaster</i> nyálmirigypár, az 1.4 ábra szerinti <i>fluoreszcens</i> markerrendszerben: DsRed magenta, GFP zöld és DAPI kék színnel ábrázolva, először színenként, majd egyesítve. A háttér fekete marad, mivel a minta megvilágítása mindig más hullámhosszon történik mint amilyen színben a jelölőanyagok fluoreszkálnak	11
1.6. A <i>GlueFlux</i> jelzőrendszer működése. Az idő előrehaladtával a fehér színű granulumok száma csökken, és egyre több magentává válik, ezt a jelenséget a granulumok beltartalmának elsavasodása okozza. A középső sorban sematikus ábrák, az alsó sorban pedig valódi mikroszkópos felvételek láthatóak. A kép forrása Dósa Anna 2022-es biológia TDK előadása [9], némi átalakítással	12
1.7. Magenta-zöld-kék és vörös-zöld-kék színezés összehasonlítása. Bal és jobb oldalon ugyanaz az információ látható, azonban jobb oldalt könnyebb összehasonlítani az első két csatornát.	12
1.8. Apotome és szeletelés nélküli (Widefield) képkészítési technológiák összehasonlítása. A bal oldalon a nyíllal jelölt objektumok jobban kivehetőek annak köszönhetően, hogy nincsenek egyéb rétegből származó zavaró jelek. („Egy 70µm vastag szelet <i>X. laevis</i> ebihalról. Kék: sejtmag (DAPI [DP]), zöld: idegek (Alexa Fluor 488 lekövetés (tracing [TR])), sárga: dendritek (Alexa Fluor 546 biocytin visszakövetés (backtracing [BT])), vörös: aktin-szálak (Alexa Fluor 647 phalloidin immunfestés [PH]). (a) ApoTome képkészítéssel és (b) szokásos epifluoreszcens kép.” Az ábra forrása Weigel, Schild és Zeug munkája [15])	13
2.1. Összefoglaló folyamatábra a modellemet megvalósító algoritmusról (1/2)	25
2.1. Összefoglaló folyamatábra a modellemet megvalósító algoritmusról (2/2)	26
2.2. DAPI <i>fluoreszcens</i> marker intenzitása egy képen, eredeti változatban (bal felső), közvetlenül limitálva (jobb felső) és <i>Gauss-elmosás</i> (<i>gaussian blur</i>) ($\sigma = 5$) után (bal alsó) illetve azután limitálva (jobb alsó). A maszkolás során a jobb alsóhoz változatot használok fel ebből a csatornából	28

2.3. Maszkolás. Az első a beolvasott (load) felvétel, második képen a maszkolás előtti DsRed (vörös), GFP (zöld) síkok és a generált mask sík (kék), a harmadik képen pedig maszkolás utáni DsRed (vörös) és GFP (zöld) látható. A felvétel tetején lévő sejt granulumjainak nagy része nincsen benne a fókuszsíkban, így a lépés után az eltűnik.	29
2.4. Felismert granulumok vizualizációja. Az első képen a vörös illetve zöld csatornán a betöltött nyers adatok láthatóak, kék színnel pedig a felismert granulumok vannak jelölve (all sík); a második képen a vörös csatornát emelem ki; az alsó sorban a felső képek tetejének (cián keret) kinagyított változata látható. Ideális esetben kék jelölő éppen kvantifikálásra alkalmas granulumokra van elhelyezve. A második képről leolvasható, hogy a felső-középső részen sok <i>szekeciós granulum</i> nincsen fókuszban, így csak 1-2 kék jelölőt látunk azon a régióban, azonban közelről megnézve az elhelyezett jelölők éppen értékelhető granulumokat takarnak.	33
3.1. Kontroll és megismételt géncsendesítés (Iru^{RNAi}) összehasonlítása, y tengelyen a <i>pozitív</i> granulumok arányával az összeshez viszonyítva. Amennyiben az elképzelések helyesek, mind a két minta pozitivitása alacsony.	37
3.2. Megismételt kísérlet kontroll (kék) és géncsendesített (Iru^{RNAi}) (piros) felvételeinek összehasonlítása, y tengelyen a <i>pozitív</i> granulumok aránya, x tengelyen az összes granulum száma az adott felvételen.	38
3.3. Géncsendesített (Iru^{RNAi}) és különböző kontrollfelvételek összehasonlítása, y tengelyen a <i>pozitív</i> granulumok arányával az összeshez (x tengely) viszonyítva. Kék színnel a 18°C-os környezetben végrehajtott kontrollkísérlet, piros színnel a 25°C-os környezetben végrehajtott összes korábbi kontrollkísérlet közül a közeli felvételek, zöld színnel pedig a géncsendesített törzs felvételei. Az új felvételeken (kék és zöld) a granulumok magas számának a magyarázata az, hogy a gyors közbelépésemnek köszönhetően nagyobb területet felölelő felvételek kerültek mentésre.	38

3.4. Különböző hőmérsékleten végzett géncsendesített (<i>Iru^{RNAi}</i>) és kontroll felvételek összehasonlítása, y tengelyen a <i>pozitív</i> granulumok arányával az összeshez (x tengely) viszonyítva. Piros és zöld színnel a 25°C-on végzett kontroll és kísérleti csoport eredményei, míg kék és sárga színnel a 18°C-on végzett kontroll és kísérleti csoport eredményei.	38
3.5. A LampRing rendszer működése. A középső sorban sematikus ábrákkal, az alsó sorban pedig mikroszkópos felvételeken követhető nyomon a gyűrűk kialakulásának folyamata LampRing rendszer esetén [9]. A magenta szín DsRed jelre, a zöld pedig GFP jelre utal.	40
3.6. Az eredeti implementáció működése. A konklúziót az utolsó panelről lehet levonni, amelyen vörös színnel a membránok, zöld színnel pedig a gyűrűk láthatóak, a két szín keverékéből a színárnyalat alapján meg lehet mondani, hogy a narancssárga az pozitív, GFP gyűrűs granulumot, míg a vörös az negatív, GFP gyűrű nélküli granulumot jelöl.	41
3.7. A vörös és a zöld csatornán a szokásos GFP és DsRed látható, a kék csatornán azonban <i>MI</i> segítségével mesterségesen előállított gyűrűk . . .	44
A.1. Kép szorzása skalárral. A jobb oldali képen színenként az intenzitás kétszer akkora, mint a bal oldalon.	51
A.2. Képek kivonása egymásból. A fenti három képen rendre ugyanannak a felvételnek a vörös (DsRed) és kék (DAPI) csatornája, illetve a kettő különbsége látható, alatta pedig a vörös csatornán a különbség, a kéken pedig az eredeti vörös csatorna. A sejtmaggal átfedésben lévő <i>szelekciós granulumok</i> némileg kékesebbek, halványabbak. A képek szélessége nagyjából 25µm	52
A.3. DAPI <i>fluoreszcens marker</i> intenzitása egy képen, limitálás előtt és után. A kör alakú formákból sejtmagok jelenlétére következtethetünk. A különbség nem szembetűnő, de ha keresünk a bal oldali képen nagyon halvány részeket, azok a jobb oldali képen hiányoznak, továbbá a világosabb részek a jobb oldalon sokkal világosabbak. A későbbiekben be fogok mutatni szembetűnőbb különbséget eredményező módszereket is.	53

A.4. Lokális limitálás (thresholding) sokkal jobb eredményeket hoz változó kontrasztú kép esetén. A felső képen egy könyvről készült fotó látható, alatt a globális limitálás majd legalul lokális limitálás. A limitálás során a felbontás romlani látszik, de ez csak információvesztésre utal (dimenziócsökkenésre). Lokális limitálás esetén a legkevesebb a hiba. A kép forrása: skimage dokumentáció [47]	55
A.5. Egy 11x11-es méretű, $\sigma = 3$ paraméterű <i>Gauss-kernel</i> és annak $y = 0$ vonal menti metszete	56
A.6. Egy kép és annak <i>Gauss-elmosás</i> (<i>gaussian blur</i>) utáni változata. A harmadik képen a vörös csatornán az eredeti, a zöld csatornán pedig az elmosott változat látható, a képet jobban megfigyelve feltűnhet, hogy a granulumok nagy része sárga (piros + zöld), csak az átlagosnál nagyobb intenzitású granulumok kapnak pirosas árnyalatot.	56
A.7. A <i>Difference of Gaussians</i> (<i>DoG</i>) művelet szemléltetése. Egy kép eredeti, elmosott változata majd az eredeti és az elmosott különbsége. Alatta a különbség abszolút értéke látható, amelyen minden lokális elváltozás kirajzolódik. Legalul a vörös csatornán az eredetiből vonom ki az elmosottat, míg a zöld csatornán fordítva, így a vörös csatornán a lokális <u>intenzitásnövekedések</u> , míg a zöld csatornán a lokális <u>intenzitáscsökkenések</u> rajzolódnak ki.	57
A.8. Az <i>erózió</i> és a <i>diláció</i> . A második ábrán az első ábra erodált változata, míg a harmadik ábrán a dilált változata látható, 5px sugarú kör alakú kernellel.	60
A.9. Az opening és a closing. A második ábrán az első ábra opening utáni változata, míg a harmadik ábrán a closing utáni változata látható, 5px sugarú kör alakú kernellel.	61
B.1. A letölthető program használata	64
B.2. Képernyőkép a fő program felhasználói felületéről	65
B.3. A felhasználói felület állapotai. A diagramon kékkel jelöltem azokat az állapotokat, amelyek megadhatóak akkor, ha a címsort átadjuk valakinek, esetleg elmentjük az oldalt könyvjelzőbe (1/2)	66

B.3. A felhasználói felület állapotai. A diagramon kékkel jelöltem azokat az állapotokat, amelyek megadhatóak akkor, ha a címsort átadjuk valakinek, esetleg elmentjük az oldalt könyvjelzőbe (2/2)	67
B.4. Néhány képernyőkép <i>dashboard</i> komponens felhasználói felületéről . .	68

Szójegyzék

Drosophila melanogaster Közismert nevén ecetmuslica, előszeretettel használt modellállat biológiai kutatásokban a gyors reprodukciós rátájának és alacsony fenntartási költségeinek köszönhetően. 5, 7, 11, 78–80, 86

dashboard A program egy elkülönülő komponense, amely alkalmas nagy mennyiségű kvantifikációs adat vizualizálására, esetleg a fő programmal való kommunikációra. 24, 31, 32, 46, 63, 68, 69, 71, 84

dekonvolúció A mikroszkópos képek zajának és torzításának csökkentése. Minden lencsének van egy torzítása és zaja. A leggyakrabban használt fluoreszcens mikroszkópok képesek nem csak felülnézeti vetületről, hanem akár mélyebb rétegekből való képalkotásra is. Szerencsére a mikroszkóp torzítása invertálható, a zajszűrésre azonban bonyolultabb heurisztikák szükségesek, amelyek nem képzik tárgyát ennek a dolgozatnak. 19

diláció Morfológiai művelet. Teljes *kernel* mentén végzett minimum-keresés. Megfelel a minkowski-különbségnek [51]. 59, 60, 83

DoG Difference of Gaussians. 31, 45, 49, 57, 58, 83

DoH Difference of Hessians. 31, 58

ELTE ASF ELTE Biológiai intézet Anatómiai Sejt- és Fejlődésbiológiai Tanszék. 17, 36, 49

erózió Morfológiai művelet. Teljes *kernel* mentén végzett maximum-keresés. Megfelel a minkowski-összegnek [51]. 60, 83

flood-fill algoritmus Tükörfordításban elárasztó-kitöltő algoritmus. Egyszerű algoritmus a kontúrral körbezárt terület kijelölésére. Iteratív eljárás, amelynek

kiindulópontja egy olyan pontnak az egyelemű halmaza, amely benne van a kitöltendő területben, és amelyben minden iteráció során hozzávesszük a halmaz elemeinek nem-kontúr szomszédait. 42

fluoreszcens marker Olyan mesterséges jelölőanyag, ami adott hullámhosszú fénnnyel való megvilágítás hatására más hullámhosszú fényt bocsájt ki. A vizsgált képeken (jó esetben) csak ilyenek látszanak. 9–11, 14, 27, 28, 30, 31, 53, 79, 80, 82, 86

Gauss-elmosás (gaussian blur) *Gauss-kernelel történő konvolúció.* 27–29, 54, 56, 80, 83

Gauss-kernel Olyan *kernel*, aminek a vonalmenti metszetének az intenzitása a σ szóráshoz normális eloszlás eloszlásfüggvényét követi. Egy-egy elemét a következő képlet írja le: $K[x, y] = e^{-\frac{x^2+y^2}{2\sigma^2}}$. 56, 83, 86

Glue Olyan ragasztószerű, cukoroldalláncokat tartalmazó anyag összessége, amelyet a *Drosophila melanogaster* lárva a bábozódás során önmagának környezethez való rögzítéséhez használ [8] [7]. A tárgyalt biológiai kísérletben [13] a szekréció folyamán ilyen anyag kerül kiürítésre. 8–10, 86

GlueFlux jelzőrendszer Olyan *fluoreszcens markerrendszer*, amely magéntában fluoreszkáló DsRed fehérjét és a zöldben fluoreszkáló GFP fehérjét is hozzáköti a *szekréciós granulumok*ban fellelhető *Glue*hoz, így az a *Glue*, amely savas, fehér (magenta+zöld, DsRed+GFP) színben látszik, amely azonban savas, csak magéntában [13]. 7, 10, 12, 29–32, 37, 39, 46–48, 79, 80

ground truth Magyarra alapigazsággként fordítható. Statisztikai és gépi tanulási szakkifejezés, olyan információ, amiről axiómaként feltehető, hogy igaz. A képfeldolgozásban olyan felvételeket jelent, amelyek a feldolgozandó képeknél nagyságrendekkel pontosabbak. [44] Én a kifejezést a hagyományos értelemben használom, tehát olyan információkra, amelyekről egy szakértő kijelentette, hogy igazak. 35, 36

hiperparaméter Mesterséges intelligencia szakkifejezés. Olyan paraméter, amelynek értékét csak kísérletezéssel lehet meghatározni, nem számítható ki más paraméterek függvényében. 32

- kernel** Egy kisebb adatmátrix amely rávetíthető egy kép összes pixelének egy környezetére. Általában a kép széleihez közel levő pixelek speciális kezelést igényelnek, mert azok környezete nem feltétlenül része a képnek. 54, 59, 60, 85, 86
- konvolúció** A képfeldolgozás egyik legsokoldalúbb alapvető művelete. Ez egy bináris képtranszformációs művelet képek és adatmátrixok (ún. kernelek) között, amely során az összes pixelre a környezetük és az adatmátrix között skaláris szorzatot képzünk [48]. 54, 59, 86
- krinofágia** A sejtekben természetes körülmények között lezajló folyamat, amely során *lizoszómák* közvetlenül egyesülnek a kiürülésre váró *szekrétációs granulumokkal*. 7–10, 14, 46, 79
- lizoszóma** Egy membránnal határolt sejtservecske, amiben emésztőenzimek vannak. 8, 87
- LoG** Laplace of Gaussians. 31, 58
- magas szintű** Azokat a programozási nyelveket nevezzük $\sim$ nek, amelyekben egy-egy utasítás több száz-ezer gépi utasítást takar [55]. A dolgozatomban a kifejezést átvitt értelemben alkalmazom, $\sim$ modell alatt olyan modellt értek, aminek egyes részlépései hasonlóak egy magas szintű programozási nyelv egyes utasításaihoz. 21
- MI** mesterséges intelligencia. 44, 46, 82
- negatív** (a dolgozatomban) olyan *szekrétációs granulumok*, amelyek elsavasodtak, így nem tartalmazzak GFP jelet, csak DsRed-et. 5, 14, 16, 32, 34–36, 47, 78
- pipeline** Jól elkülöníthetően lépcsőzetesen egymásra épülő blokkokból álló algoritmus. 22, 87
- pipeline_lib.py** Egy önálló keretrendszer mely a munka melléktermékeként készült. Alkalmas *pipeline* algoritmusok implementálására. 22, 23, 46
- plane** Sík. Az implementációmban használt angol szó egyes eltárolt képváltozatok jelölésére. 27

pozitív (a dolgozatomban) olyan *szekeációs granulumok*, amelyek nem savasodtak el, így tartalmaznak GFP és DsRed jelet is. 5, 14, 16, 32, 34–39, 47, 78, 81, 82

RPF time Relative to Puparium Formation. 39

sarokdetektálás (blob detection) Alacsony szintű 2D-s vagy 3D-s struktúrák felismerése képeken [56]. Ez a gyakorlatban éppen átvilágított gömbök elmosott vetületére illik legjobban, amely pontosan megfelel a *szekeációs granulumok* fluoreszcens mikroszkópon keresztül látott változatának. 31, 57, 58

scale-space A kép paraméteres transzformáltjainak összessége, ahol a paraméter megfeleltethető a kép méretezésének [25]. Egy képfeldolgozási megközelítésként is értelmezhető. 31, 58

szakértői modell Olyan modell (bizonyos adatok alapján bizonyos predikcióra alkalmas komplex (bonyolult) függvény / algoritmus), amelynek az előállítása szakértőktől származó információk alapján történt. 6, 9, 15, 17, 24, 32, 46–48

szekeációs granulum Membránnal határolt váladékszemcse. 4–8, 10, 14, 16, 18, 29, 30, 33, 47, 48, 52, 78, 79, 81, 82, 86–88

trichotóm összehasonlítás Két valós szám között a $<$, $>$, $=$ relációk közül pontosan egy áll fenn, a $\sim$ ezek közül a megfelelő kiválasztása [57]. 30

unió Két halmaz egyesítése. Például az $\{1,2\}$ és $\{2,3\}$ halmazok $\sim$ ja $\{1,2,3\}$. Matematikai jelekkel: $\{1,2\} \cup \{2,3\} = \{1,2,3\}$. 16, 31