



Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Távközlési és Médiainformatikai Tanszék

Szív MRI felvételek előállítása generatív mélytanuló eljárásokkal

TDK dolgozat

Készítette:

Meglécz Máté

Konzulens:

Kalapos András
dr. Gyires-Tóth Bálint

2023

Tartalomjegyzék

Kivonat	i
Abstract	ii
1. Bevezetés	1
2. Elméleti és irodalmi háttér	3
2.1. Mélytanuló eljárások	3
2.2. Generatív modellek	4
2.3. Generatív versengő hálózat	4
2.3.1. Általános felépítése	4
2.3.2. GAN alkalmazásai egészségügyi felvételek generálására	5
2.4. Denoising Diffusional Probabilistic model	7
2.4.1. Általános felépítése	7
2.4.2. Semantic diffusion model	8
2.4.3. Diffúziós modellek alkalmazásai egészségügyi felvételek generálására	8
3. Adathalmaz bemutatása	11
4. Rendszerterv	14
5. Kiértékelési módszerek bemutatása	16
5.1. Fréchet Inception Distance	16
5.2. Kernel Inception Distance	16
6. GAN módszer	18
6.1. Modell és implementáció	18
6.2. Tanítások	18
7. SDM módszer	23
7.1. Modell és implementáció	23
7.2. Tanítások	23
8. A megoldások értékelése	28
9. A módszer alkalmazási lehetőségei	30
10. Összefoglalás	32

Köszönetnyilvánítás	34
Irodalomjegyzék	35
Függelék	37
F.1. T1 apex	37
F.2. T1 mid	38
F.3. T1 base	39
F.4. T2 apex	40
F.5. T2 mid	41
F.6. T2 base	42

Kivonat

A szív működését, a szívizom eltéréseit különböző módon lehet vizsgálni. Ezek közül napjaink egyik legfejlettebb nem invazív módszere a CMR (kardiovaszkuláris mágneses rezonancia) képalkotás. A CMR vizsgálat egyik fő előnye, hogy szöveti információt is nyújt, melynek megítéléséhez újabban a szívizom esetén kvantitatív információ is nyerhető, az ún. T1 és T2 térképezésen alapuló módszer segítségével. A normál szívizomszövet meghatározott T1 és T2 relaxációs idővel rendelkezik, mely kóros állapotokban pl. szívizomelhalás, heg megváltozik. A mélytanuló eljárások számos egészségügyi témakörben jelentős sikereket értek el, azonban minden új vizsgálati módszer egyedi megoldás kidolgozását teszi szükségessé. Továbbá az eredmények kutatáson túlmutató, termékszintű felhasználása még csak kevés esetben történt meg.

Egy feladat, lehet a témakörben, hogy ezeken a szív felvételeken egy modell képes legyen automatikusan szegmentálni a bal kamrát és annak a falát. A mélytanuló eljárások minőségét bizonyos esetekben jelentősen befolyásolhatják a tanító adatok mennyisége, változatossága és minősége. Az ilyen orvosi feladatok automatizálásra szolgáló megoldások kialakításának egy nehézsége a megfelelő adathalmazhoz való hozzájutás. Egy gyakori probléma, hogy kevés adatot tartalmaz egy adathalmaz, hiszen egyfelől nem készülnek nagy mennyiségben MRI felvételek, másfelől nehéz hozzáférni, hiszen a kórházak, nem szolgáltatathatják ki könnyedén a betegekhez tartozó adatokat, illetve még ha ez lehetséges is, akkor is az adatokat megfelelően anonimizálni kell.

Ezen okok miatt, merült fel az ötlet, hogy kedvező lenne, ha rendelkeznének egy olyan eljárással, ami képes általunk létrehozott szegmentációs képekből „hamis” MRI felvételeket előállítani. Így adatdúsítást tudnánk végezni az új generált képekkel, majd az új adathalmazon történő tanításokkal megpróbálni javítani más modellek teljesítőképességét.

Napjainkban elterjedt megoldás képgenerálási feladatokra a generatív versengő hálózatok (GAN), illetve a diffúziós modellek használata. A dolgozatban ezen két módszer alkalmazása kerül bemutatásra és értékelésre szív MRI felvételek generálására a bal kamra és annak falának szegmentációja alapján. Mind a generatív versengő hálózatok, mind a diffúziós modellek magas minőségben képesek előállítani felvételeket. A diffúziós modellek jobb diverzitás lefedő képességgel rendelkeznek, míg a GAN-ok gyorsabban hajtják végre a generálási folyamatot.

A téma Városmajori Szív- és Érgyógyászati Klinika, illetve a Siemens Healthcare együttműködésével történik.

Abstract

The function of the heart and abnormalities of the myocardium can be studied in different ways. One of today's most advanced non-invasive methods is CMR (cardiovascular magnetic resonance) imaging. One of the main advantages of CMR is that it provides tissue information, and more recently quantitative information can also be obtained for the myocardium using a method based on T1 and T2 mapping. Normal myocardial tissue has a defined T1 and T2 relaxation time, which changes in abnormal conditions such as myocardial death or scarring. Deep learning techniques have achieved significant success in many health-related areas, but each new examination method requires a unique solution. Furthermore, there have been only a few cases where the results have been applied beyond research to the product level.

One task could be to use these heart recordings to automatically segment the left ventricle and its wall. In some cases, the quality of deep learning procedures can be significantly affected by the quantity, variety and quality of the training data. One difficulty in developing solutions to automate such medical tasks is access to an appropriate data set. A common problem is that there is little data in a dataset, since on the one hand MRI images are not produced in large quantities, and on the other hand they are difficult to access, since hospitals cannot easily provide the data associated with patients, and even if they can, the data must be properly anonymised.

For these reasons, the idea arose that it would be beneficial to have a procedure that could produce "fake" MRI images from the segmentation maps we created. In this way, we could perform data enrichment with the new generated images and then try to improve the performance of other models by training on the new data set.

Nowadays, a common solution for image generation tasks is to use generative adversarial networks (GANs) or diffusion models. In this paper, the application of these two methods to generate cardiac MRI images based on segmentation of the left ventricle and its wall is presented and evaluated. Both generative adversarial networks and diffusion models can produce high quality images. Diffusion models have better diversity coverage, while GANs perform the sampling process faster.

The topic is carried out in cooperation with Városmajor Heart and Vascular Center and Siemens Healthcare.

1. fejezet

Bevezetés

Az utóbbi években egyre nagyobb figyelmet kapnak a gépi tanuló, azon belül is a mélytanuló eljárások. Ez a kiemelt figyelem betudható, mind a hardwarek fejlődésének, mind az eljárások fejlődésének. Egyik legfontosabb alkalmazásuk folyamatok automatizálásában van. Az egyik fontos iparág, ahol folyamatok automatizálásával esetenként jelentős minőségbeli javulást lehet elérni, az egészségügy.

Egészségügyben mesterséges intelligencia által asszisztált rendszerek, képesek segíteni az orvosok munkáját, azzal például, hogy repetitív munkafolyamatokat automatizálnak ki. Ezzel nem csak időt spórolnak az orvosnak, hanem akár az ellátást is megbízhatóbbá teszik, hiszen az emberi hiba, mulasztás valószínűsége lecsökken. Természetesen ezek a rendszerek sem működnek hibátlanul, viszont segítségükkel az orvosoknak bizonyos folyamatokat csak felügyelnie kell, mint szakértő. Amennyiben szükséges az orvos rendelkezik a szükséges tudással, hogy közbelépjen vagy felülbírálja a mesterséges intelligencia által hozott döntéseket. Mint döntéstamogató rendszerek meg tudják könnyíteni a diagnózisok felállítását, azzal hogy feltudnak fedezni olyan összefüggéseket a páciensek leleteiben, amiket egyébként nem biztos, hogy egy ember észre venne. Ezek mellett fontos előnye az egészségügyben történő automatizációnak, hogy akár az ellátásban az orvos több időt tudjon egy pácienssel foglalkozni.

Az egyik ilyen automatizálható munkafolyamat a szív MRI (Magnetic Resonance Imaging, mágnesrezonancia-képkalkotás) felvételeken a bal kamra és annak a falának a szegmentálása. Ezeknek a körbe jelölése egy hosszabb folyamat a kardiológusok számára is. Egy ilyen feladatra megoldásként szolgálhatnak a képszegmentáló neurális hálózatok. Viszont ahhoz, hogy egy ilyen modellt jól belehessen tanítani, nagy méretű adathalmazra van szükség, ahol az MRI felvételek rendelkeznek a szükséges szegmentálási maszkokkal.

Az egészségügyi adathalmazokra sajnos jellemző, hogy nem rendelkeznek nagy mennyiségű adatokkal. Ez két okból adódik, egyfelől MRI felvételek nem készülnek minden nap egy emberről, másfelől ha készülnek is efféle felvételek, akkor a személyiségi jogokra való tekintettel nehéz hozzájuk férni. Mielőtt egy ilyen adathalmaz felhasználhatóvá válik, anonimizáláson kell átesnie.

Erre a problémára megoldásként szolgálhatnak a generatív mélytanuló eljárások. Segítségükkel realiztikus adatokat lehet generálni, amikkel fel lehet dúsítani az eredeti adathalmazokat. Ezen új adathalmazon tanítva más modelleket azt várhatjuk, hogy azok teljesítménye javulni fog a megnövekedett adatmennyiség hatására.

Ebben a dolgozatban realiztikusnak tűnő szív MRI felvételek generálása a cél, a szív bal kamrájának és annak a falának a szegmentációja alapján, hogy a későbbiekben ezekkel a maszkkal rendelkező felvételekkel, már létező adathalmazok méretét lehessen növelni és később az új adathalmazra tanítani más modelleket.

Szív bal kamra és annak falához tartozó szegmentációs maszk alapján történő, szív MRI felvétel generálásra az irodalomban lehet találni példát generatív versengő hálózat (GAN, Generative, Adversarial Networks) által. Viszont ezen feladatra a dolgozatom írásának pillanatában diffúziós modellen alapuló megoldás nem létezik. Ahhoz, hogy a diffúziós módszer összehasonlítható legyen az irodalomban elterjedten alkalmazott GAN módszerrel, mindkét modellt szükséges betanítanom az általam birtokolt adathalmazra, mivel az irodalomban más adathalmazokat használtak a modellek tanítására.

A dolgozatban először rövid elméleti bevezetőt adok a későbbiekben bemutatott két generatív módszerhez, majd egyesével áttekintem az ezekhez és a feladatom jellegéhez illeszkedő irodalmat. Ezek után bemutatom az adathalmazom felépítését, majd részletesen bemutatom először a generatív versengő hálózattal az eredményeimet, majd a szemantikus diffúziós modellel is hasonlóan járok el. Végezetül a dolgozatom végén összehasonlítom a két eljárást.

2. fejezet

Elméleti és irodalmi háttér

Ebben a fejezetben a munkámhoz kapcsolódó elméleti és irodalmi háttereket mutatom be. Először egy rövid elméleti bevezetőt adok a dolgozatban későbbiekben alkalmazott mélytanuló eljárásokhoz, majd áttekintem a két módszer hasonló témákban publikált irodalmait.

2.1. Mélytanuló eljárások

A különböző mélytanuló eljárások az utóbbi években nagy figyelmet kapnak több felhasználási területen is, azonban a témám a képfeldolgozás és adatdúsítás területére szűkíthető be, így elsősorban ezen témák bemutatására törekszem.

A konvolúciós neurális hálózatok (CNNs, Convolutional Neural Networks) [17] state-of-the-art megoldásnak tekinthetők számos területen, mint például objektum észleléses, kép szegmentációs és klasszifikációs feladatok esetén. Az eljárás alapját a konvolúciós rétegek adják, amik számos szűrő (filter) segítségével képesek kinyerni, esetemben képekből, olyan jellemvonásokat, amik jelentősen elősegítik a teljesen összekötött rétegeknek a tanulást. A konvolúció művelete grafikus kártyákon (GPU, graphics processing unit) kiválóan párhuzamosítható, ami szintén hozzájárul az eljárás előnyeihez. Feladatomhoz két fontos architektúra emelendő ki, mivel tanításaim során ezeket alkalmaztam.

Az első ilyen architektúra a U-Net [21], ami eredetileg szegmentációs feladatra lett kifejlesztve. A hálózat egy encoder és egy decoder útvonalból áll, ami az u-alakú felépítést adja. Az encoder hálózat végzi a jellemzők kinyerését és megtanulja bemenet egy absztrakt reprezentációját. A decoder hálózat pedig ebből az absztrakt reprezentációból állítja elő a képhez tartozó szegmentációt. A két ágot egy híd köti össze néhány konvolúciós réteggel.

A másik architektúra a ResNet [8], residual blokkokból áll néhány up/down-sampling réteg között. Alapvetően a elenyésző gradiens (vanishing gradient) probléma megoldását szolgálja nagy modellek esetén azzal, hogy egy identitás összeköttetés (identity connection) segítségével egy blokk bemenetét hozzáadja a blokk kimenetéhez, így a blokk belseje egy „residual mapping”-et tanul meg, ami lényegében a kimenet és a bemenet különbségét adja.

2.2. Generatív modellek

Generatív modellek a gépi tanulás egy olyan része, ahol a feladat a tanító adatoknak egy leképzését megtanulni, hogy azzal új adatokat lehessen létrehozni. Általában mély neurális hálózatokra épülnek.

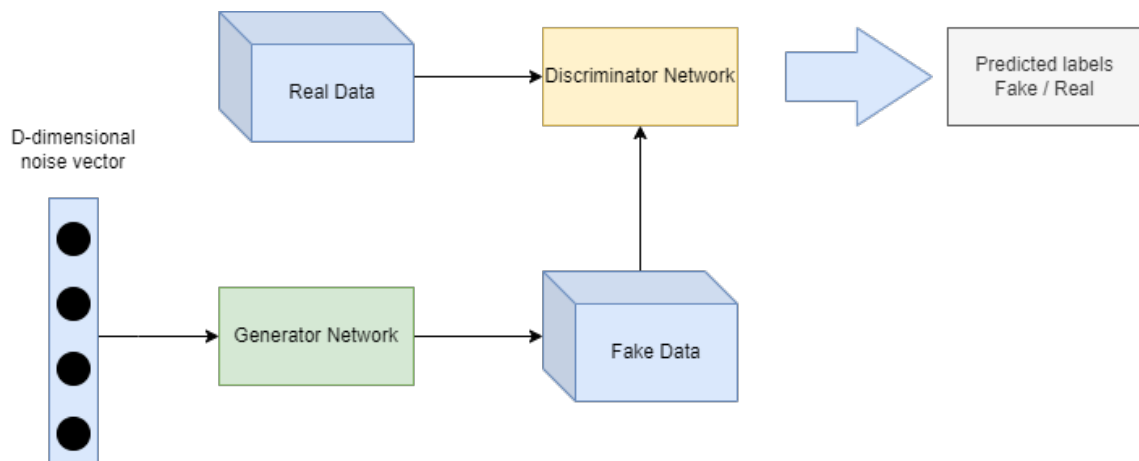
Erősségük, hogy segítségével realiztikus "hamis" adatokat lehet generálni. Egyik jellemző alkalmazása az adathűtés, jellemzően kisméretű adathalmazok esetén. Szöveg generálási és művészeti feladatokban is széleskörűen lehet alkalmazni.

A 3 legelterjedtebb módszer képgenerálásra a generatív versengő hálózat [7], a diffúziós modell [25] és a variációs autoenkóder [12]. A dolgozat további része az első kettőre fókuszál, mivel a variációs autoenkóderek jellemzően gyengébb minőségű adatok generálására képesek.

2.3. Generatív versengő hálózat

2.3.1. Általános felépítése

Feladatomban új felvételek generálására az egyik megoldásként generative adversarial hálózatot [7] tanítottam. GAN modelleknek számos változata létezik, itt csak egy általános ismertetőt adok, az alkalmazott hálózat későbbi fejezetben kerül bemutatásra. A GAN-ok 2 részmodellből épülnek fel, egy generátorból és egy diszkriminátorból. A GAN általános felépítését a 2.1 ábrán lehet megtekinteni. A generátor modellt arra tanítjuk, hogy új adatokat generáljon, a diszkriminátort pedig, hogy eldöntse, hogy a bemenete a generátor által előállított adat vagy igazi adat a tanító adathalmazból. A két részmodell a tanítás során folyamatosan verseng egymással, a generátor célja, hogy becsapja a diszkriminátort, a diszkriminátort pedig, hogy felismerje a generált adatokat.



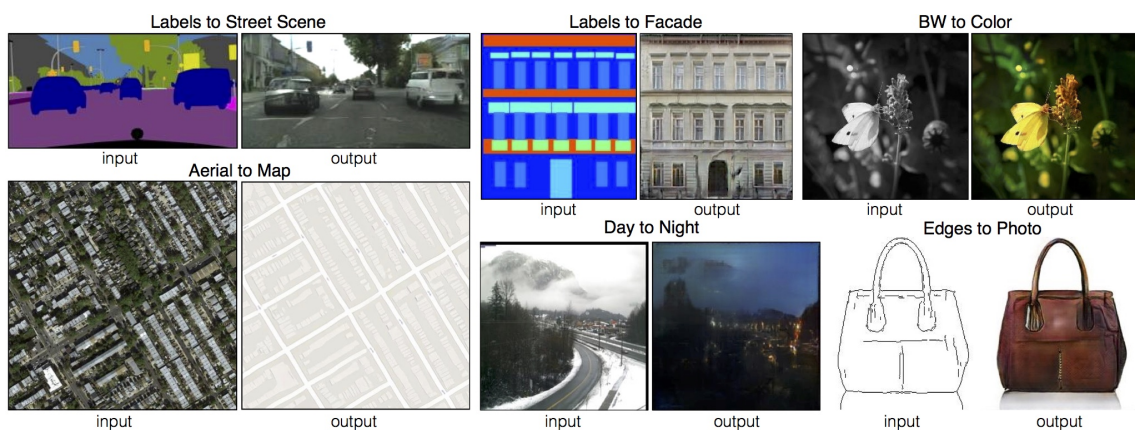
2.1. ábra. GAN általános felépítése. (az ábra [15] alapján módosítva)

2.3.2. GAN alkalmazásai egészségügyi felvételek generálására

Irodalomkutatásom elsősorban a képgeneráló, "image-to-image" GAN-ok és ezek egészségügyi képképzésben való felhasználására terjedt ki. A publikált munkák között azt tapasztaltam, hogy könnyebben lehetett agyi MRI felvételekről szóló írásokat találni, azonban ezekből az alkalmazásokból származó megoldások és tapasztalatok is hasznosak, hiszen ezek az MRI felvételek is hasonlóan komplexek, mint a szívről készültek. Az egyik első cikk [1] is egy áttekintést ad az agyi képgenerálás tekintetében, összefoglalja az addig publikált írásokat több szemszögből is, esetemben az alkalmazott GAN variánsok és az alkalmazott kiértékelési technikák voltak értékes információk.

Ezután találtam egy másik áttekintő cikket [23], viszont ez már a kardiológiában felhasználható GAN modellekről foglalkozott. Bemutatja a publikus adathalmazokat, ezek jellemzőit, viszont mint korábban, ebből a cikkből is elsősorban a felhasznált GAN modellekről szóló részek segítettek abban, hogy megismerjem, hogy milyen lehetőségeim vannak a megfelelő modell kiválasztására.

A korábbi tapasztalatokból kiindulva egyértelművé vált, hogy egy conditional GAN [14] kategóriába sorolható modellre lesz szükségem. A conditional GAN (cGAN, feltételes GAN), olyan generatív modell, ahol a generátor és a diszkriminátor is kap plusz kényszert. Esetemben ilyen információ a bal kamra és a falának a szegmentációja, egyfelől a generátornak olyan képet kell generálnia, ami illeszkedik a szegmentációra, másfelől a diszkriminátornak az is feladata, hogy megítélje, hogy a kép mellett, hogy igaznak tűnik-e, illeszkedik-e a szegmentációhoz. A korábban említett cikkekben több ilyen típusú GAN is említésre került, ezek közül kiemelendők az alábbiak: Pix2PixGAN [10], SPADE GAN [18], StyleGAN [11], Cycle GAN [28].



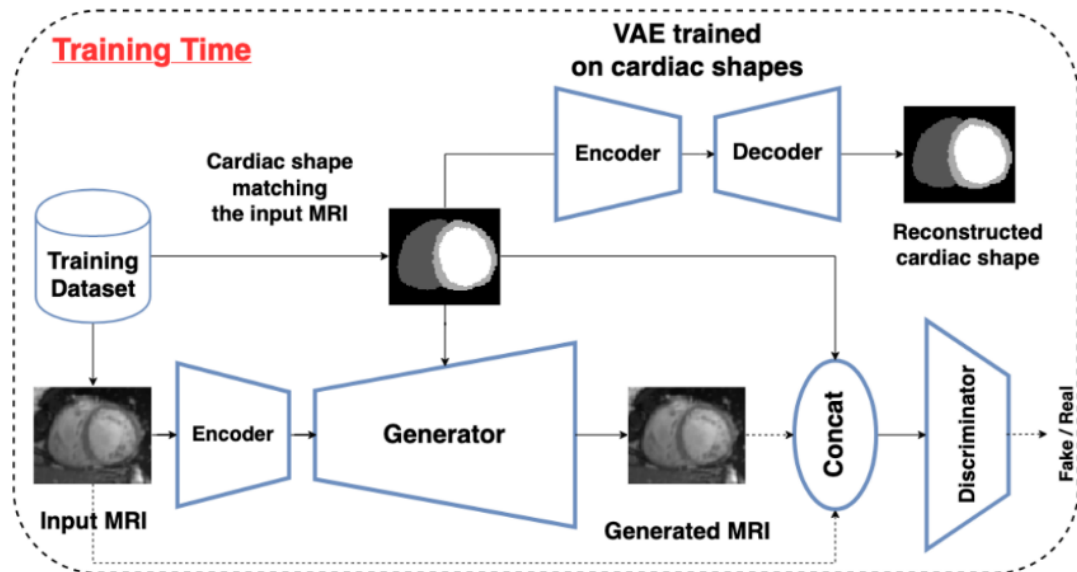
2.2. ábra. Pix2PixGAN alkalmazásának példái. (az ábra forrása: [10])

A Pix2Pix GAN [10] egy bemeneti képből generál egy másik képet, pixeleket rendel pixelekhez. Az eredeti cikk példái között, mint a 2.2 ábrán látszik, olyan alkalmazások szerepelnek, ahol a bemeneti kép szegmentáció, kontúr kép, az igazi kép egy módosított változata (más időjárási körülmények). A generátor architektúrája a korábban bemutatásra került U-Net architektúra.

Egy másik cikkben [22] is Pix2Pix modelleket használnak. Itt szintetikus tumorral rendelkező agyi MRI felvételeket szeretnének generálni publikus adathalmazok felhasználásával. 3 modellt is betanítanak, az első modell egy „image-to-brain” szegmentációt tanul, ahol az MRI felvételekből szegmentációk kerülnek kigenerálásra. A második modell „label-to-image” szintézist csinál, ahol a szegmentációkból generálnak MRI felvétel-szerű adatokat. Az utolsó modell pedig „image-to-tumor”, ahol MRI felvételekről generálják ki a tumor szegmentációját. Ezekre azért van szükség, mert 2 adathalmazt egyesítenek és az egyikben csak agy szegmentációval, míg a másikon csak tumor szegmentációval rendelkeznek. Végül adat szintézises megoldásukban először legenerálják minden MRI felvétel agy szegmentációját, majd ezeken elhelyeznek tumor szegmentációkat. Végezetül ezekből kigenerálják a szintetikus MRI felvételeket.

A SPADEGAN [18] a Pix2Pix modell egy továbbfejlesztett változata, bevezet egy Spatially Adaptive Normalization réteget, ami segíti a szemantikus információ megtartását a hálón keresztül. Hagyományos normalizációs, konvolúciós és nem lineáris rétegekkel ezek az információk „elmosódhatnak” a tanítás során.

Ennek alkalmazását szív felvételek generálására találtam. A cikkben [24] a GAN mellett egy variációs autoenkóder is tanítanak, hogy a szegmentációkat egy látens térből lehessen mintavételezni. Ebben az alkalmazásban a generátor nem csak a szegmentációkat kapja meg információként, hanem bemenetként kap egy „stílus” képet is, amihez hasonlónak kell a generált képnek is lennie. A cikkben bemutatott megoldás felépítését a 2.3 ábra mutatja be.



2.3. ábra. MRI felvételek generálása SPADEGAN-nal úgy, hogy a szegmentációs maszkok variációs autoenkóder segítségével generálódnak. (az ábra forrása: [24])

StyleGAN [11] modellekkel a generált képek általános stílusát lehet kontrollálni, mint például: fényviszonyok, emberek esetén például hajszín, bőrszín.

A CycleGAN [28] is hasonlóan stílus alapján generálja a képeket, az eredeti cikkben példaként olyan alkalmazásokat hoznak, ahol egy fénykép bemenetként érkezik és egy festő stílusában generálódik belőle a kimenet. Ennek a modellnek pél-

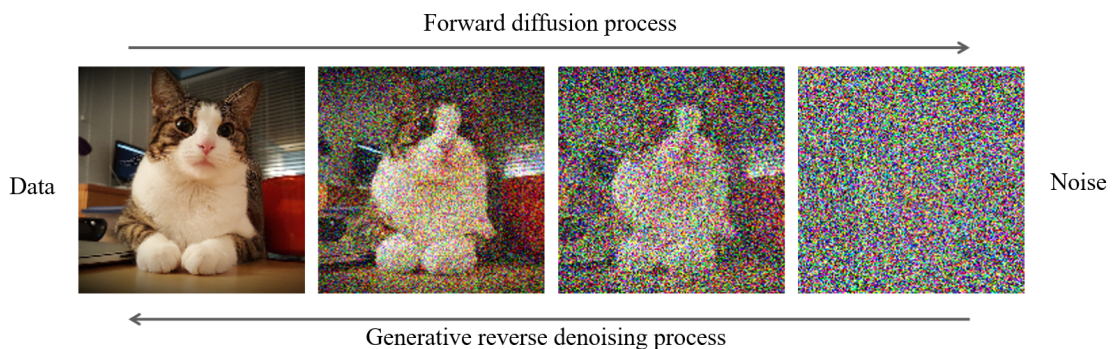
dául olyan felhasználását találtam az orvosi képgenerálás témában, ahol agyi MRI felvételekből generálnak CT (computed tomography, komputertomográfia) felvételeket [20].

A korábban említett két modell feladatomból szempontjából kevésbé alkalmazható, hiszen nem a stílusát kell változtatni az MRI felvételeknek, hanem szegmentációkból generálni őket.

2.4. Denoising Diffusional Probabilistic model

Diffúziós modellek [25] napjainkban az egyik state of the art megoldásnak tekinthetők képgenerálási folyamatokban. Feladata egy eloszlást Markov-lánc segítségével egy másik eloszlássá transzformálni. Két fő részből épül fel, a forward és a backward diffúziós folyamatból.

2.4.1. Általános felépítése



2.4. ábra. A forward és backward process demonstrálása DDPM-ben. (az ábra forrása: [16])

A forward process-ben fokozatosan egyre több zajt ad a képhez, egészen addig, amíg teljesen felismerhetetlenné válik. A folyamat során a tanító adathalmaz komplex és ismeretlen eloszlását egy egyszerűbb eloszlássá alakítja.

A backward process az előző folyamat által létrehozott „korruptált” képből indul ki. A célja lépésről lépésre eltávolítani a zajt a képről, hogy a végén minél jobban az eredeti eloszlást kapja vissza.

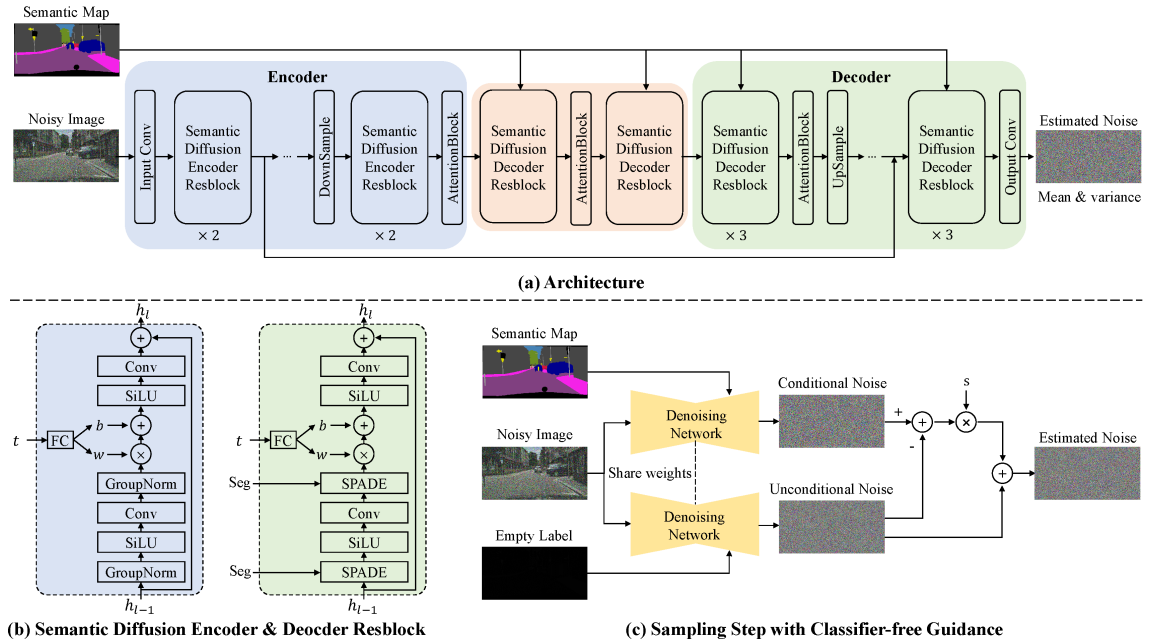
A forward folyamat során a lépések valószínűségi sűrűség függvényei (PDF, Probability Distribution Function) eltérnek egymástól. A backward folyamat során egy mély tanuló megoldást használunk, hogy meghatározzuk a forward lépések PDF paramétereit. A két folyamat működését a 2.4 mutatja be.

Ez a megoldás jelentősen stabilabb tanítás és mintavételezés közben, mint más megközelítések szintetizálási feladatokhoz, mint például a GAN vagy a VAE (variational autoencoder, variációs autoenkóder).

2.4.2. Semantic diffusion model

A Semantic Diffusion Model (SDM, szemantikus diffúziós modell) [27] a DDPM-ek (Denoising Diffusion Probabilistic Model, zajtalanító diffúziós valószínűségi modell) között egy olyan modell, ami lehetőséget ad egy maszk bemenettel befolyásolni a generált képet. Korábbi conditional megoldásokban a maszk és a bemeneti kép együtt került beadásra az enkóder U-net architektúrájának, azonban ez nem mindig adott a maszknak megfelelő kimenetet.

Ebben a modellben egy olyan megoldást kínálnak, ahol a maszkot csak a backward folyamatban kapja meg a modell SPADE [18] rétegeken keresztül. Ezek mellett alkalmaz a mintavételezésre egy úgynevezett classifier free guidance stratégiát, ami során nem csak a maszkkal generált képet veszi figyelembe, hanem az anélkül generáltat is. A kettőből hozza létre a végleges kimenetet. A megoldás felépítését a 2.5 ábra mutatja be. A cikkben összehasonlítják a modell teljesítményét más elterjedt generatív megoldásokkal is. A bemutatott megoldás majdnem minden esetben jobban teljesít, mint más elterjedt módszerek.



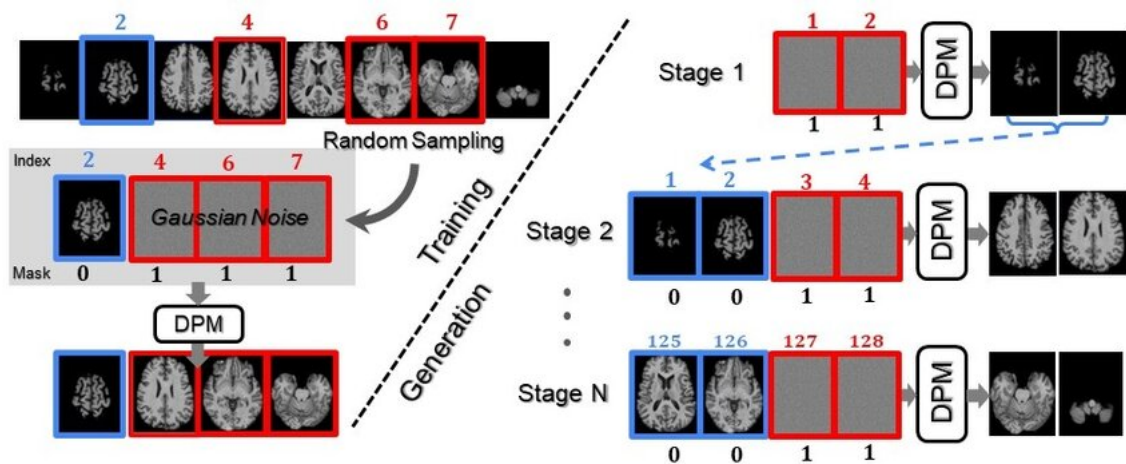
2.5. ábra. (a) A feltételes denoising hálózat architektúrája. A denoising hálózat bemenetként a zajos képet veszi, és megbecsüli a szemantikus maszk alapján. (b) Az enkóder és dekóder resblock részletes szerkezete. (c) A mintavételezéses classifier free guidance eljárással. (az ábra forrása: [27])

2.4.3. Diffúziós modellek alkalmazásai egészségügyi felvételek generálására

Diffúziós modellek területén is az irodalomkutatásom elsősorban olyan cikkekre fókuszált, amik orvosi témában készültek és olyan modelleket használnak, amikben a generált felvételek kondicionálhatóak egy adott bemenetre. Ezen bemenetek jellemzően maszkok vagy kontúrtérképek.

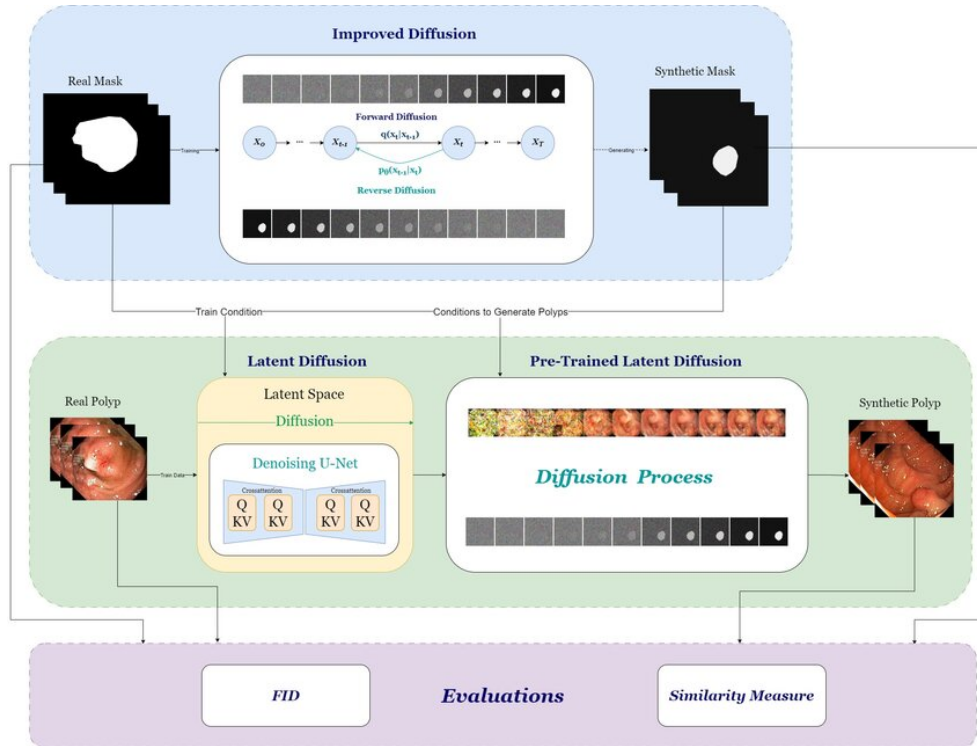
Egyes cikkek 3 dimenziós agyi MRI felvételek szintetizálását célozzák meg. Med-DDPM [6] a 3D-DDPM-t [5] fejleszti tovább, azzal hogy a modell a bemenetén vár egy maszkot is, amivel meg lehet határozni a tumor elhelyezkedését. A szerzők kitérnek a modell egyik legfontosabb erősségére is, hogy segítségével könnyedén lehet anonimizált adathalmazokat generálni, ezzel megoldva az egészségügyi adatoknál felmerülő titoktartási (privacy) problémákat is.

Egy másik munkában [19] szintén agyi MRI felvételeken dolgoznak, azonban nem az egész 3D felvétel szintetizálása a cél. Az agyi metszetek generálására egy olyan megoldást mutatnak be a szerzők (2.6 ábra), ahol a tanítása során néhány példányt kiválaszt a metszet felvételekből és azokat feltételes bemenetként használva generálja a hiányzó részeket. Mintavételezés során pedig zajból generálja az első néhány metszetet, majd ezeket használja fel a maradék felvétel generálásához.



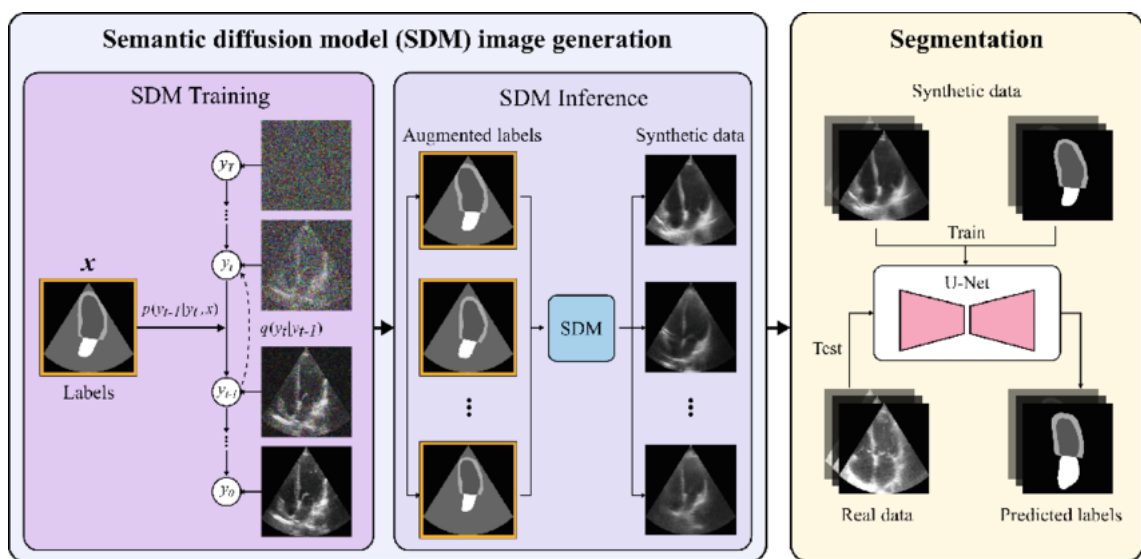
2.6. ábra. A tanítás és a mintavételezés agyi metszetekre. (az ábra forrása: [19])

Az irodalomban nem csak MRI felvételek generálására lehet példákat találni. Gasztrointesztinális polip felvételek maszk alapú szintetizálására [13] is lehetséges példát találni. A cikkben a felvételek szintetizálása mellett, a maszkok generálását is diffúziós modellekkel oldják meg (2.7 ábra). Ezzel még tovább egyszerűsítik az adatszűrés folyamatát, hiszen a maszkokat sem emberi munkával kell létrehozni.



2.7. ábra. Polip felvételek generálása. (az ábra forrása: [13])

A cikk [26] Semantic diffusion model segítségével szív ultrahang felvételeket generál szegmentációs maszkok alapján. Bemutatja, hogy a diffúziós megoldással jobb minőségű felvételeket lehet generálni, mint versengő hálózatokkal. A cikkben bemutatott megoldást a 2.8 ábra szemlélteti. Ezen megoldásnál fontos megemlíteni, hogy az implementáció is publikusan elérhető, ezzel könnyítve a bemutatott eredmények reprodukálhatóságát. Ezen cikk a publikus implementáció és a feladat hasonlósága, miatt fontos kiindulási alapként szolgált az általam elvégzett munkához.



2.8. ábra. Szív ultrahangok generálása SDM-mel. (az ábra forrása: [26])

3. fejezet

Adathalmaz bemutatása

Az adathalmazomban 223 beteg T1 és T2 szívfelvételei találhatóak meg a bal kamra és annak falának szegmentációjával. Az MRI felvételek előzetesen anonimizálva lettek, így védve a betegek személyiségi jogait.

Ezek a felvételek DICOM (Digital Imaging and COmmunications in Medicine, Digitális képkezelés és kommunikáció az orvostudományban) fájlok, amikben 1 csatornán és 16 biten van kódolva a színmélység. A DICOM egy olyan szabvány, ami orvosi kép alkotásában a felvételek tárolását, kezelését, küldését és nyomtatását határozza meg. Segítségével a különböző képalkotó eljárásokkal (MRI, CT, röntgen, ...) készített felvételeket egységesen lehet kezelni. A felvételek tárolása mellett a fájlok metaadatai között a páciensről és a képalkotás eljárásáról találhatóak információk. Egy kontúr térkép tartozik a felvételekhez, ami a bal kamrát és annak falát jelöli ki. Ezekből a kontúr térképekből lehetséges szegmentációs maszkokat generálni.

A T1 és T2 típusú felvételeken belül 3 típus különíthető el, az apex, a mid és a base (lásd 3.2 és 3.3 ábrák). A T2 relaxációs idő az az idő, amely alatt a szövetben lévő protonok visszatérnek egyensúlyi állapotukba a keresztirányú síkban. T1 relaxációs idő esetén ez az idő hosszanti síkra vonatkozik. A T1 típusú felvételeken a zsírszövetek világosabbak lesznek, míg a folyadékok sötétebbek, segítségével például azonosítható és megkülönböztethető a szív különböző struktúrái, mint például a szívizom és a környező szövetek a szívben. A T2 felvételeken a zsírszövetek és a vízbázisú szövetek kerülnek kiemelésre. T2 felvételek segítenek azonosítani a szívizom ödéma területeit, amelyek különféle szívbetegségekre utalhatnak, mint például a szívizomgyulladás.

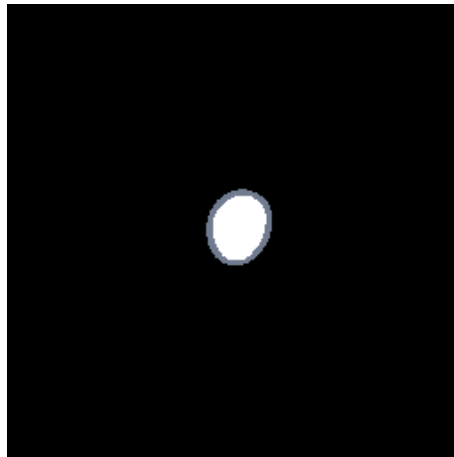
Az apex, a mid és a base a szív különböző mélységben alkotott képmetszeteit különböztetik meg. Az apex felvételek a szív csúcsi (alsó) részéről, a mid felvételek a középső részről, míg a base felvételek a szív bazális (felső) részéről készülnek. Ezen felvételek segítségével lehet például megvizsgálni a kamrák állapotát, a szív és a szívbillentyűk helyes működését. Egy-egy típusú felvételhez egy páciens esetében több felvétel is rendelkezhet ugyanazzal a szegmentációs maszkkal (lásd 3.1 ábra).

Az összes beteg felvételeiből minden tanítás során 25 beteg adatai elkülönítésre kerültek, ezek a képek később tesztelésre és a betanított modellek kiértékelésére lettek felhasználva. Így nagyjából 6200 felvételt tudtam a tanításra használni a maradék alany képei közül. A modell tesztelésére pedig 960 felvétel állt rendelkezésre a 25 páciens szívéből.

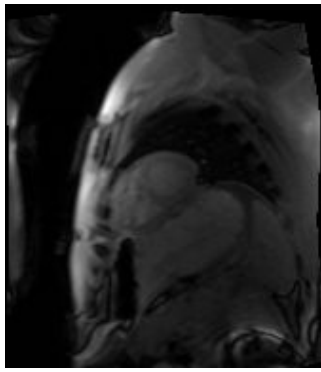
A 3.1 táblázatból látható, hogy az adathalmaz nem kiegyensúlyozott, T1 felvételekből kétszer annyit tartalmaz, mint T2 felvételekből. Összeségében a felvételek így is elégé hasonlítanak egymásra, a tanításokhoz ezt nem ítéltém problémának.

3.1. táblázat. Az adathalmazban a különböző típusú adatok megoszlása.

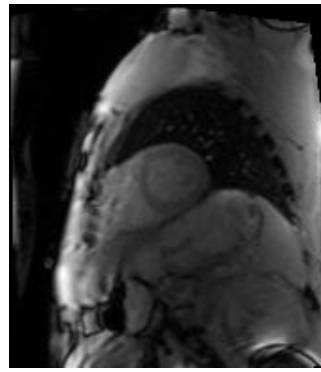
Felvétel típusa	Mennyisége a tanító adathalmazban	Mennyisége a teszt adathalmazban
T1 Apex	1185	215
T1 Base	1550	215
T1 Mid	1426	215
T1 felvételek száma összesen	4061	675
T2 Apex	681	86
T2 Base	731	86
T2 Mid	734	86
T2 felvételek száma összesen	2146	288



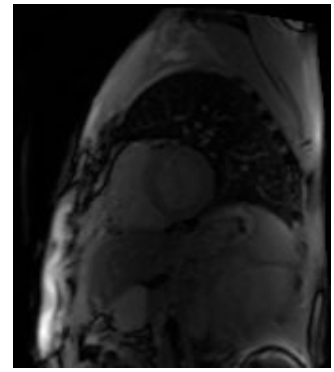
3.1. ábra. Példa egy szegmentációs maszkra.



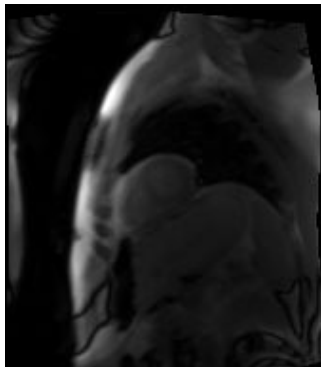
(a) T1 apex



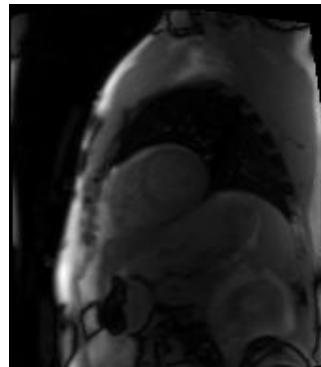
(b) T1 mid



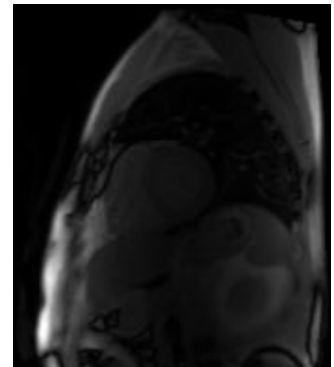
(c) T1 base



(d) T2 apex

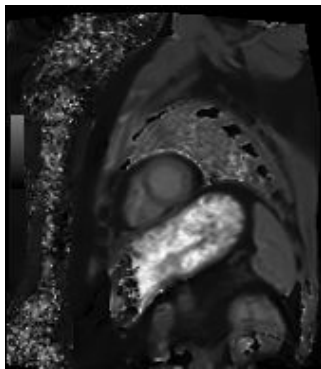


(e) T2 mid

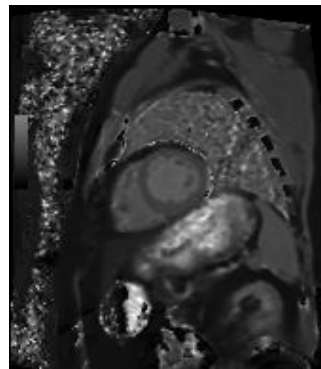


(f) T2 base

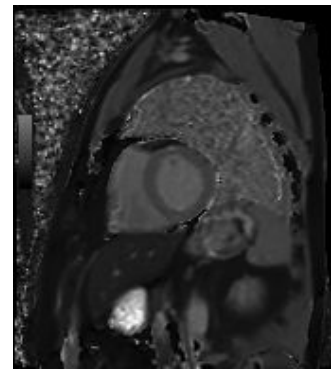
3.2. ábra. Különböző típusú felvételek



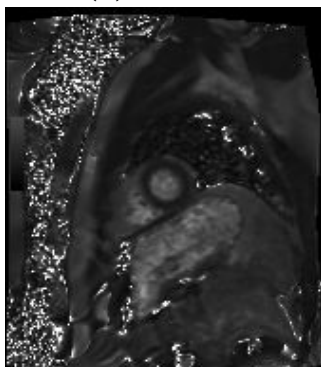
(a) T1 apex



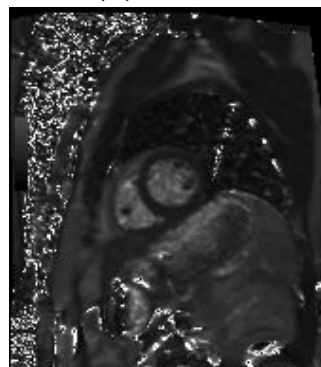
(b) T1 mid



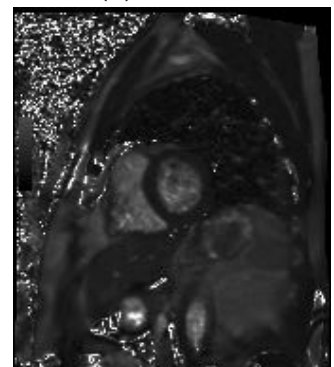
(c) T1 base



(d) T2 apex



(e) T2 mid



(f) T2 base

3.3. ábra. Különböző típusú mapping felvételek

4. fejezet

Rendszerterv

Ebben a fejezetben a későbbi alkalmazáshoz a rendszertervet mutatom be. A rendszer elsődleges célja bemeneti maszkokból, realiztikus felvételeket generálni, hogy azokkal adathalmazokat lehessen dúsítani.

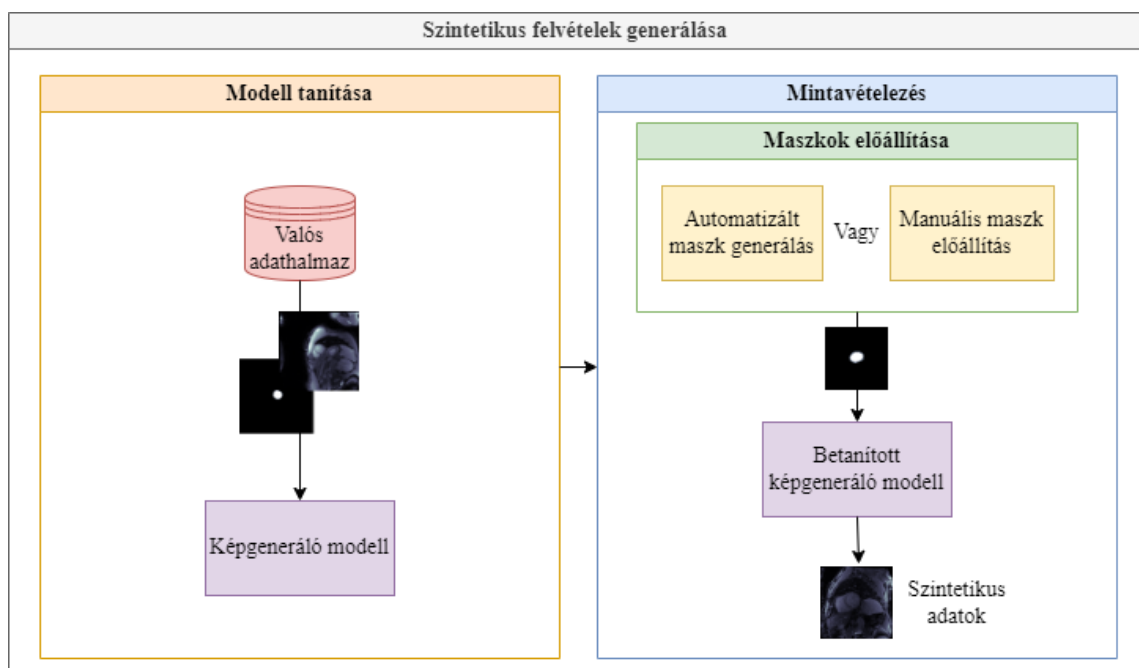
Először a rendszerben egy már létező adathalmaz felhasználásával egy képgeneráló modell betanításra kerül, hogy a korábban definiált feladatot képes legyen ellátni.

Miután a modell sikeresen betanításra került, a mintavételezés fázisa következik. A mintavételezés során maszkokból a modell előállítja a szintetikus felvételeket. A maszkok előállításának két lehetséges módszere létezik.

Egyfelől manuálisan, emberi munkával létre lehet hozni ezeket, így befolyásolva, a generált felvételeket. Másfelől ez is egy automatizálható feladat, mint az irodalomkutatás során bemutatott cikkekből látni lehetett, [24] a GAN tanítása mellett egy VAE-t is tanított, hogy a maszkokat előállítsa. [13] pedig két diffúziós modellt tanított, egyet a maszkok, egyet pedig a polip felvételek generálására maszkok alapján.

Végezetül a generált szintetikus felvételek, áteshetnek még valamilyen minőség ellenőrzésen vagy közvetlenül bekerülhetnek az új adathalmazba. A generált felvételek minőség ellenőrzésére, akár egy megoldás lehet a GAN diszkriminátorához hasonló modell, ami eldönti, hogy elég realiztikus-e a szintetikus felvétel.

A rendszertervet a 4.1 ábra mutatja be. A dolgozatomban, csak a modellek betanításával és már meglévő maszkok alapján történő mintavételezéssel foglalkozik.



4.1. ábra. Rendszerterv

5. fejezet

Kiértékelési módszerek bemutatása

A generatív módszerek által generált képeket elsőre nem egyértelmű, hogy hogyan lehet objektívan kiértékelni, hiszen nehéz kifejezni azt metrikákkal, hogy ember számára mennyire elfogadható, valószínű a generált kép. Több cikkben előfordult, hogy szubjektívan, orvosok eldöntötték képekről, hogy szerintük az adott kép igazi vagy hamis, generált kép. Ilyenre nekem a munkám során nem volt lehetőségem.

Objektív mérésneként gyakran láttam használni a Fréchet Inception Distance-t és a Kernel Inception Distance (KID). Teszteléskor mindkét mérőszám kiszámítása megtörténik a teszt adathalmazon. A TorchMetrics python könyvtárban implementálva vannak ezek, így annak segítségével értékeltem ki a generált képeket.

5.1. Fréchet Inception Distance

A FID [9] metrika az InceptionV3 mély konvolúciós neurális hálózatot használja, amely egy nagy képi adathalmazon (pl. ImageNet) van előtanítva. Ezt a hálózatot jellemző kinyerőként (feature extractor) használja a generált és a valós képek statisztikáinak kiszámítására egy jellemző térben.

A FID pontszámot két statisztika alapján számítják ki: a valós és a generált képek jellemzőreprezentációinak átlaga és kovarianciája alapján. E statisztikák összehasonlításával a FID mérőszám azt számszerűsíti, hogy a generált képek eloszlása mennyire felel meg a valós képek eloszlásának. Az alacsonyabb FID pontszám azt jelzi, hogy a generált képek közelebb állnak a valós képeloszláshoz, ami magasabb minőséget és változatosságot jelent.

A FID figyelembe veszi mind az egyes generált képek minőségét, mind a teljes generált adathalmaz diverzitását. A generatív modellezés területén széles körben használt értékelési metrikává vált, mivel jól korrelál a képminőség emberi érzékelésével.

5.2. Kernel Inception Distance

A KID [4] metrika az Inception Score (IS) [3] mérőszámon alapul, amelyet eredetileg a generált képek minőségének és változatosságának mérésére vezettek be. Az Inception

Score a képek osztályozásához a FID-hez hasonlóan egy Inception modellt használ, amelyet az ImageNet nevű nagy adathalmazon tanítottak be.

A KID metrika az Inception Score-t bővíti azzal, hogy a képek osztályozási címkéi helyett a képek jellemzőinek reprezentációit veszi figyelembe. A mérőszám a valós és a generált képek jellemző reprezentációinak statisztikai közötti távolságot számítja ki. Pontosabban, kiszámítja a jellemzőreprezentációk közötti maximális átlagos eltérést (maximum mean discrepancy, MMD), amely a két eloszlás közötti különbséget számszerűsíti.

Az alacsonyabb KID érték azt jelzi, hogy a generált képek a jellemzőreprezentációk tekintetében jobban hasonlítanak a valós képekhez, ami magasabb minőségre és realizmusra utal. Ezzel szemben a magasabb KID érték nagyobb eltérést jelent a valós és a generált képek eloszlásai között.

6. fejezet

GAN módszer

Elsőként egy "baseline" megoldás kialakítására törekedtem, amihez a GAN módszert használtam fel. A GAN-nak előnye más generatív módszerekkel szemben, hogy gyorsabban lehet előállítani vele új adatokat, mint diffúziós modellekkel és jobb minőségű adatokat állít elő, mint egy Variational Autoenkoder. Diffúziós modellekkel szemben hátránya, hogy ha nagy diverzitás van az adathalmazban, azt nehezebben tudja lefedni, viszont tanítása lehetséges gyengébb hardware-n is, mert a modellek mérete jellemzően lényegesen kisebb, mint egy diffúziós modell memória igénye.

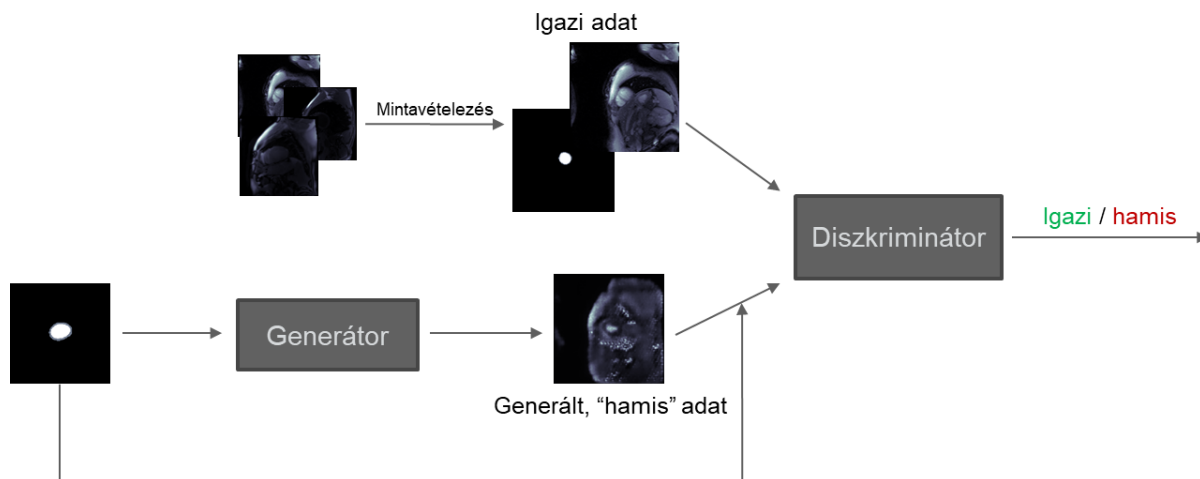
6.1. Modell és implementáció

Az irodalomkutatás során felderítettem, hogy milyen modellek használatára van lehetőségem. Egyértelművé vált, hogy célszerű lenne vagy a SPADE-nek vagy a Pix2Pix modellnek egy jól felhasználható implementációját találnom és azzal tovább dolgoznom. Jellemzően ezekhez a modellekhez publikusan elérhetőek felhasználható implementációk, viszont arra kevésbé lehet számítani, hogy aktívan foglalkoznának ezekkel az implementációkkal. A Pix2Pix modell rendelkezett olyan publikus implementációval, ami működőképes volt, így azzal foglalkoztam. A GAN felépítését a feladatomra a 6.1 ábra mutatja be.

A projektet át kellett alakítani, hogy képes legyen fogadni a dicom kiterjesztésű felvételeket, illetve azokat megfelelően meg tudja jeleníteni kiértékelés során. Ezek mellett naplózást implementáltam bele, hogy felügyelni tudjam a tanítások állapotát.

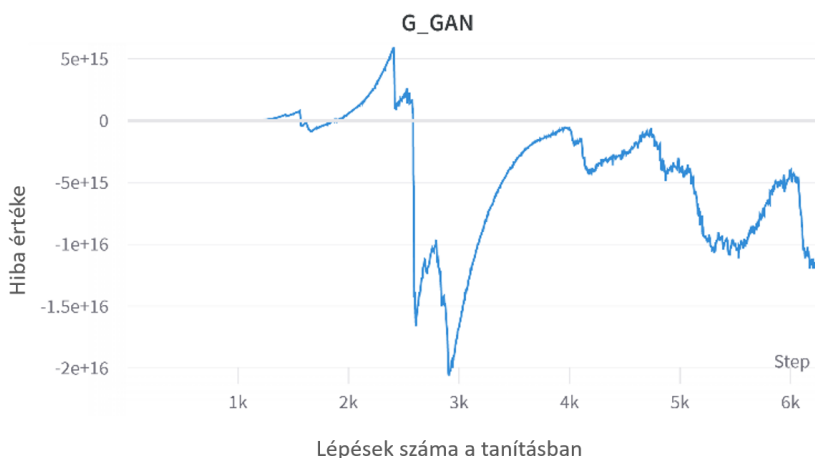
6.2. Tanítások

Az első tanítások során az alapértelmezett beállításokkal igyekeztem tanítani és az első cél egy olyan batch méret belövése volt, amivel még pont beleférnek a 12 GB memóriámban. A batch méreteket jellemzően 2 hatványokként választjuk meg, valójában ennek nagy jelentősége nincs, viszont így is végül a 128-as batch méret bizonyult megfelelőnek. Innen igyekeztem mindig ezzel a batch mérettel dolgozni, hiszen a hyperparaméterek optimalizálása során gyakori, hogy ha egyet megváltoztatunk akkor sok másikat is meg kell. Mivel az optimalizálás főleg empirikus úton történik érdemes néhányat lefixálni és a későbbiekben csak akkor hozzányúlni, ha arra mindenképpen szükség van.



6.1. ábra. GAN felépítése a feladatra.

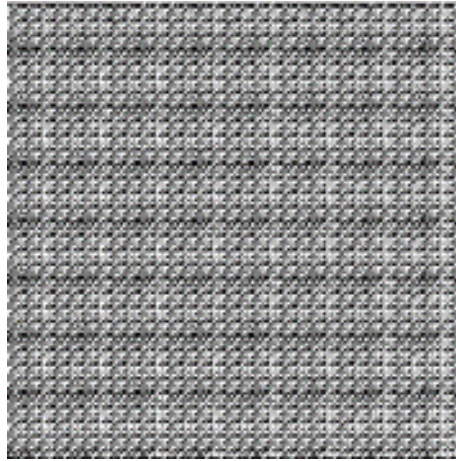
A kezdeti tanítások során hamar szembesültem azzal, hogy a tanítás során a generátor és a diszkriminátor veszteség értékei sok információval nem szolgálnak, azon kívül, hogy nagyjából az az ideális, ha minél kevésbé változnak. A GAN-ról általánosságban kijelenthető, hogy elég instabil tud lenni tanítás közben (lásd 6.2 ábra). Amikor a generátor veszteség értékei elkezdenek hirtelen növekedni/csökkeni, akkor általában tisztán zaj kezd generálódni (lásd 6.3 ábra) és ebből a helyzetből nagyon ritka, hogy újra stabilá válik a tanítás.



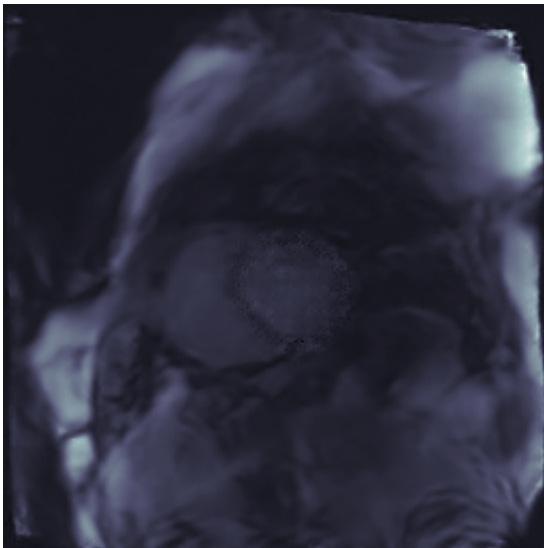
6.2. ábra. Instabillá váló tanítás.

Az első sikeresnek tűnő tanításokat az alapértelmezett beállításokkal értem el. A generátor az U-Net architektúrát követte és 1000 epoch-ig tanítottam úgy, hogy a learnig rate lineárisan csökkent minden epoch-ban 0.001-ről. A tanítás során kimenetelt képek ígéretesnek tűntek, azonban a teszt adathalmazon generált képek, már nem hasonlítottak az eredeti felvételekre (lásd 6.4 ábra). Viszont az ezeken keletkezett FID és KID értékek felállítottak egy baseline-t a további kiértékelésekre. FID: 3.64, KID (szórás, átlag): 0.0063, 0.7357.

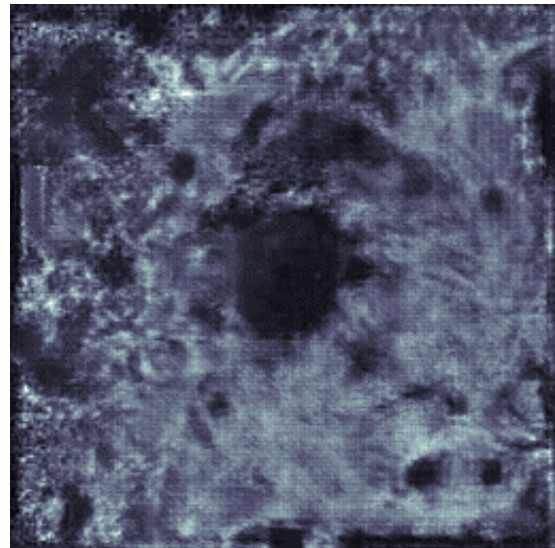
Ekkor még semmilyen augmentációt nem implementáltam a képekhez. Ilyen MRI felvételek esetén több augmentációs technikát rögtön elvettem. Ilyen volt például a képek tükrözése, hiszen a bal kamra nagyjából fix helyeken található a



6.3. ábra. Generált "zaj" felvétel.



(a) Tanítás közben generált felvétel



(b) Kiértékelés közben generált felvétel

6.4. ábra. Az első "működő" modellek teljesítménye

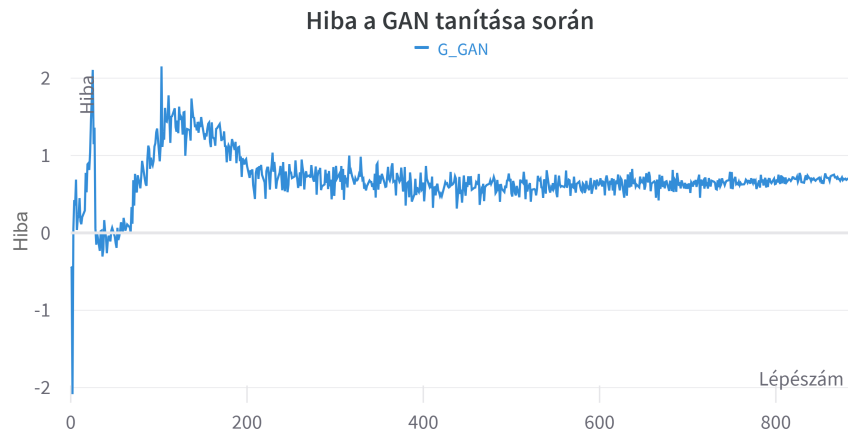
felvételeken és nem lenne hiteles, ha a bal kamra például jobb oldalra kerülne. A képek orientációja is kötött, így a forgatást sem tartottam elfogadhatónak. Az egyetlen augmentáció, amit implementáltam, a képrészletek kivágása volt, amitől eredetileg a kamra zajos generálásának a javulását vártam. Összeségében nem tapasztaltam javulást és romlást sem az augmentáció hatására, így innentől a legtöbb tanításban véletlenszerűen alkalmaztam a képek egyharmadán.

Az első olyan tanítások, amikben a teszt adathalmaz is már ígéretes volt a korábban említett architektúrákból keletkezett az epoch szám és learning rate állítgatásával, viszont ugyanazon beállítások futtatása mellett nem volt reprodukálható a tanítás. Amikor ResNet architektúrát próbáltam használni a generátorban, akkor is időnként születtek jobb eredmények viszont a reprodukálhatóság itt is problémát okozott.

Ezek után a loss függvény változtatásával kezdtem el kísérletezni, először a loss 11 részének súlyozásával igyekeztem elősegíteni, hogy a generált kép jobban

hasonlítson az eredeti képre. Az l1 loss a bemeneti és a kimeneti képek közötti eltérést bünteti. Látványos javulást azonban nem sikerült elérnem vele.

Később áttértem a wasserstein loss-ra [2], amivel sok alkalmazási területen jobb eredményeket lehetett elérni, mint a klasszikus GAN veszteségfüggvénnyel. Kezdetben elég instabil volt ezzel a tanítás, viszont amikor sikerült megtalálni a megfelelő learning rate-t, akkor lényeges javulást sikerült elérnem a generált képek minőségében. Míg a „vanilla” loss függvénnyel a legjobb elért KID átlag értékeim 0.2933 voltak, addig az új loss függvény használásával ez az érték rögtön 0.2 körülire csökkent, emellett a tanítások sokkal inkább reprodukálhatóvá váltak.



6.5. ábra. GAN sikeres tanítása során a hibafüggvény.

A GAN-nal végzett munkám vége felé még a diszkriminátor réteg számának állításával, és mind a generátorban, mind a diszkriminátorban a filterek számának beállításával kísérleteztem még. Viszont ezen tanításoknak a hátránya volt, hogy mivel növeltem a hálókbán a súlyok számát, megnőtt a memória igényük és így csökkentenem kellett a batch méretét, hogy beférjek a GPU memóriájába. Sajnos ezen paraméterek állításával nem sikerült jobb eredményeket elérnem, mint a korábbi legjobb tanítás.

Végezetül még végeztem az eddigi legjobb beállítással tanításokat úgy, hogy a T1 felvételekre és a T2 felvételekre külön-külön tanítottam. Itt sem tapasztaltam, hogy ettől jobbak lettek volna a generált felvételek az eddigi legjobb tanításhoz képest.

A legjobb tanítást 128-as batch mérettel, U-Net architektúrával, wasserstein és felsúlyozott l1 veszteségfüggvénnyel és augmentációval sikerült elérni 400 epoch alatt. Ehhez a FID értéke 1.36, míg a KID értékei (szórás, átlag) 0.0056, 0.2755 lettek. A modell által generált példák a 6.6 ábrán és a függelékben láthatóak.



6.6. ábra. Legjobb modell által generált képek a kiértékelés során

7. fejezet

SDM módszer

A GAN módszerrel kialakított megoldás után célszerűnek láttam egy diffúziós modellen alapuló megoldás kipróbálását is. A GAN-nal generált felvételekről többnyire, meg lehet állapítani, akár nem hozzáértő ember által is, hogy nem igazi felvételek. Jellemzően homályosabbak és esetenként zajosabbak is, mint lenniük kéne. A diffúziós modellek biztatató alternatívának tűntek az irodalomkutatás során bemutatott eredményi miatt.

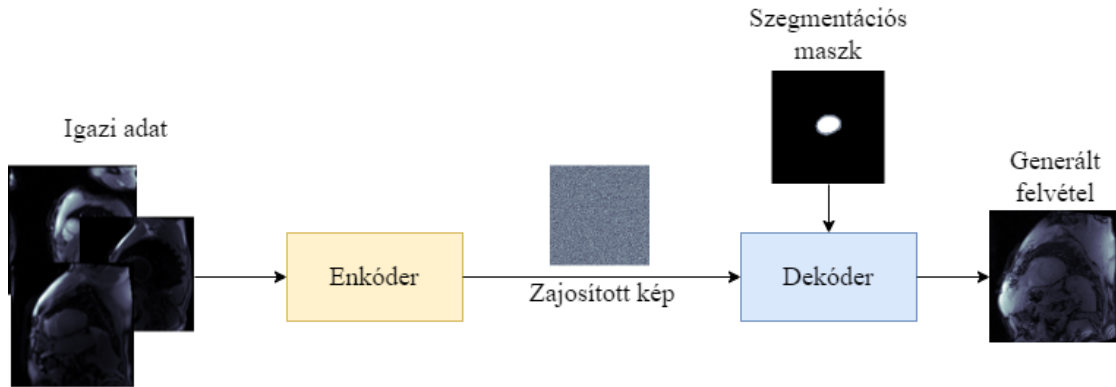
7.1. Modell és implementáció

Az irodalomkutatás során bemutattam számos megoldást hasonló feladatokra. Ezen megoldások közül kiemelkedett a szív ultrahangok generálása szegmentációs maszkok alapján. Munkám során ezt a cikket vettem segítségül két ok miatt. Egyfelől a feladat jellege sok tekintetben hasonlított az általam elvégzendő munkához, másfelől rendelkezett publikusan elérhető, működő implementációval. Ezen okok miatt Semantic diffusion model-t használtam a tanításaimra. A megoldás felépítése a feladatomra a 7.1 ábrán látható.

A GAN implementációhoz hasonlóan itt is ki kellett egészítenem a projektet, hogy a dicom kiterjesztésű felvételeimet fogadni és megjeleníteni tudja. A naplózást is implementálnom kellett ebbe a projektet is. Viszont ennél a projektnél egyéb módosításokat is végeztem, hogy kicsit jobban illeszkedjen a feladatomhoz. Alapértelmezésként 3 csatornán RGB képeket várt a modell, viszont ez nekem feleslegesnek bizonyult, mivel az MRI felvételek 1 csatornán vannak színekódolva. Ezért kiegészítettem olyan opcióval is a modellt, hogy képes legyen szürkeárnyalatos bemenetek fogadására. Ez memória használat szempontjából is előnyös, hiszen csak a bemenetek és kimenetek mérete harmadakkora így.

7.2. Tanítások

A GAN-nál lehetséges volt a saját 12 GB videó memóriával rendelkező videokártyámon tanítani, viszont az SDM esetén még, ha az architektúra méretét jelentősen is lecsökkentettem, akkor se tudtam egyáltalán beolvasni a modellt a videó memóriába. Így rögtön egy tanszéki szerveren kezdtem a tanításokat, ahol kicsit több hardware kapacitásom volt, 16 GB videó memória. Ez már elégnek bizonyult, hogy legalább el

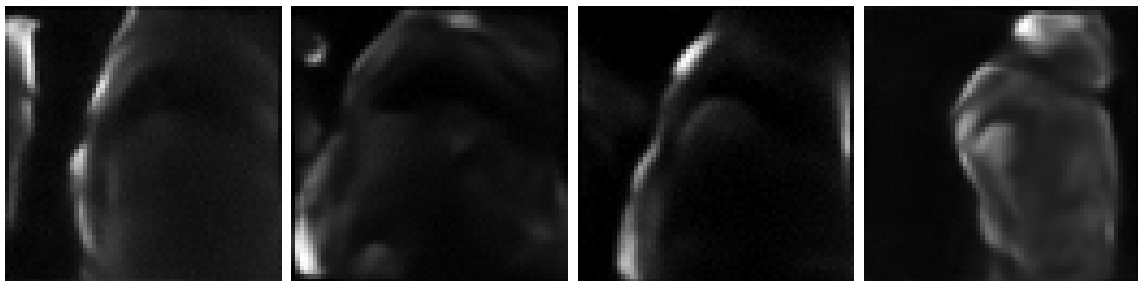


7.1. ábra. SDM felépítése a feladatra.

tudjak indítani tanításokat, bár az alapértelmezett modellt még mindig nem tudtam használni.

Kezdetben azzal próbálkoztam, hogy a resnet block-ok számát 1-re csökkentetem, így 2-es batch mérettel tudtam dolgozni. Ekkor még a képek méretét 256x256-os felbontáson hagytam, viszont feltehetőleg elsősorban a kis batchméret miatt nem igazán tudott megtanulni olyan súlyokat a háló, amivel zajon kívül mást lehetett volna generálni.

Az első ígéretes tanításokat 64x64-es felbontásra csökkentett képekkel tudtam elérni. A resnet block-ok számát még mindig lecsökkentve kellett tartanom, viszont a batch méretet így fel tudtam venni 8-ra. Ez elég sokat segített a modellnek a tanulásban. Tanítás során a felvételek már elkezdtek hasonlítani az igazi képekhez (lásd 7.2 ábra).

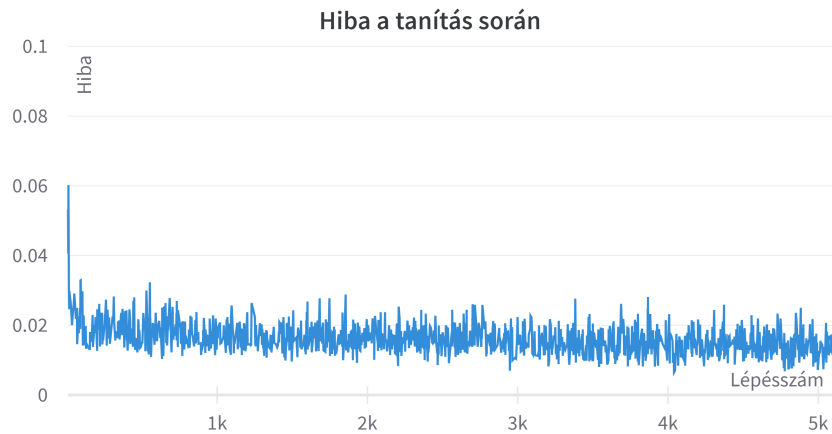


7.2. ábra. Az első "működő" sdm modell teljesítménye

Ezen a hardwaren valószínűleg ezek voltak az elérhető legjobb eredmények. Munkám folytatásához szükségessé vált még erősebb hardware. A tanszéken hozzáférést kaptam 40 GB videó memóriával rendelkező kátyákhoz, amik jelentősen megkönnyítették a munkám előrehaladását. Innentől már lényegesen kevesebb hardware által okozott korlátba ütköztem.

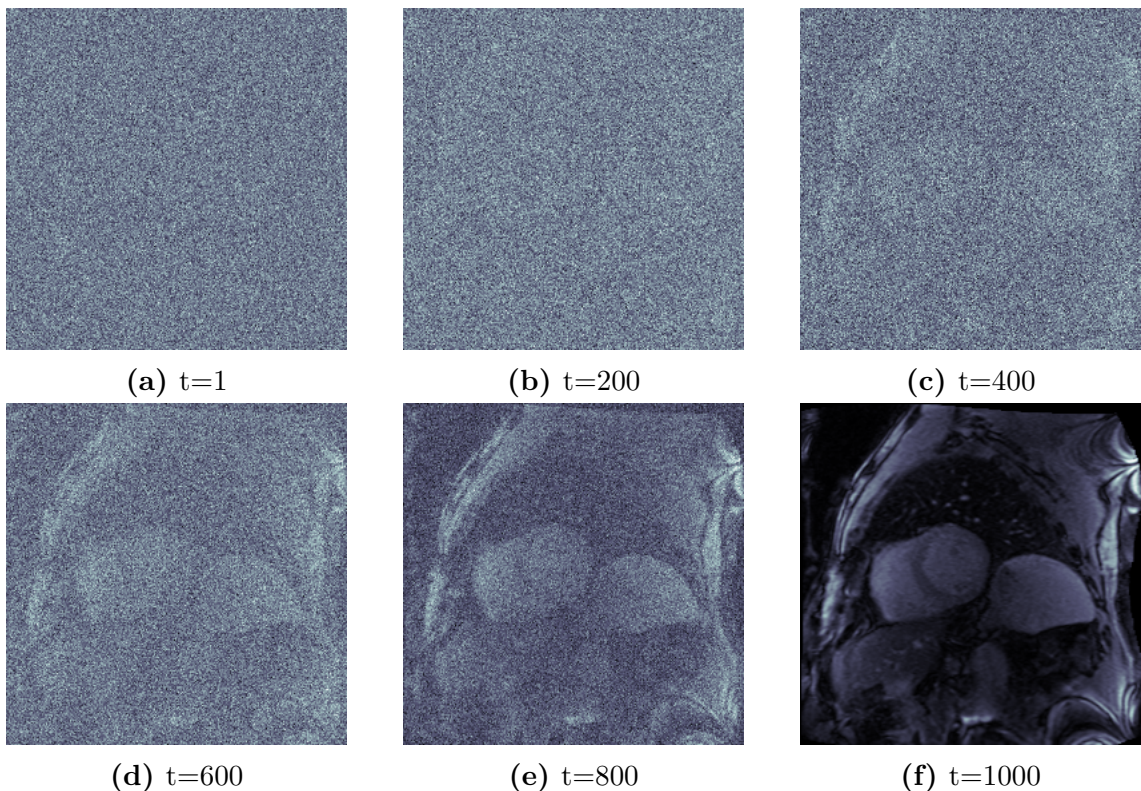
El tudtam kezdeni kihasználni a teljes modell potenciálját. Mostmár az alapértelmezett modell belefért a memóriába, úgy hogy a 256x256-os felbontást sem kellett csökkentem, a 8-10-es batch méretet és az 1000-es diffúziós lépésszámot is megtudtam tartani a tanítás során.

Stabilitási problémákat, mint a GAN-nál nem tapasztaltam, ami jelentősen megkönnyítette a tanítás folyamatát. Megfelelő méretű modellel már viszonylag könnyen tudtam realiztikus felvételeket generálni. Az első jónak tűnő eredménye-



7.3. ábra. SDM tanítása során a hibafüggvény a finomhangolás előtt.

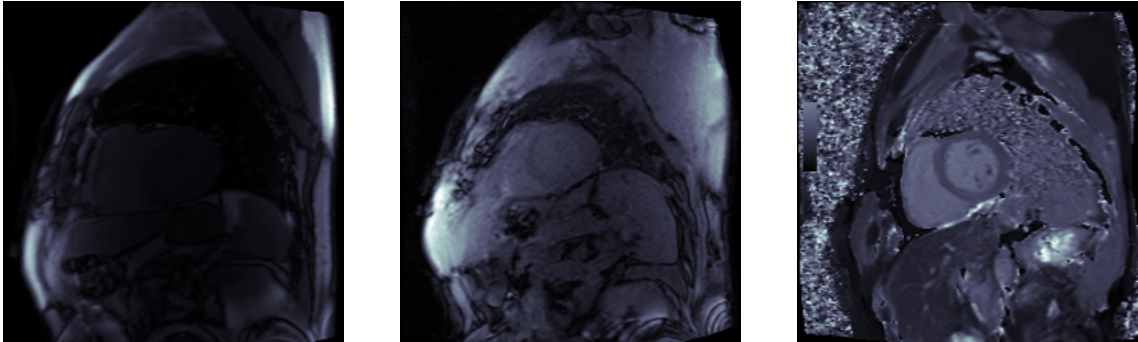
ket kiértékelve rögtön azt lehetett tapasztalni, hogy jobban teljesít, mint a GAN, mind a kiértékelési metrikákban, mind ránézésre. A FID értéke 0.65, míg a KID értékei (szórás, átlag) 0.0041 és 0.0896 lettek, amikből egyértelműen látszik a javulás a GAN-hoz képest.



7.4. ábra. A dekódolási folyamatban a felvétel állapota az egyes időpillanatokban.

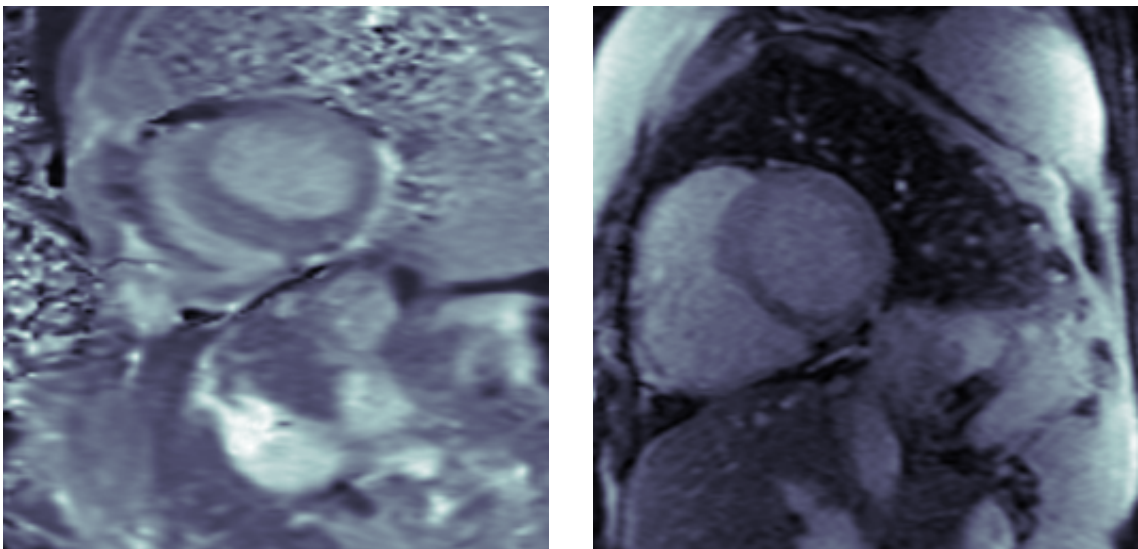
Ezt a modellt, még finomhangoltam az eredeti SDM cikkben írt technikával, 20%-os drop arány alkalmazásával. Ilyenkor a tanítás során elhagyja a maszkot a

kimenet generálásához, feltétel nélkül alkot képeket. Ezzel sikerült tovább javítanom a kiértékelési metrikákat, FID 0.45-re, míg a KID szórása 0.0036-ra, átlaga 0.0383-ra csökkent.



7.5. ábra. A finomhangolt modell által generált felvételek

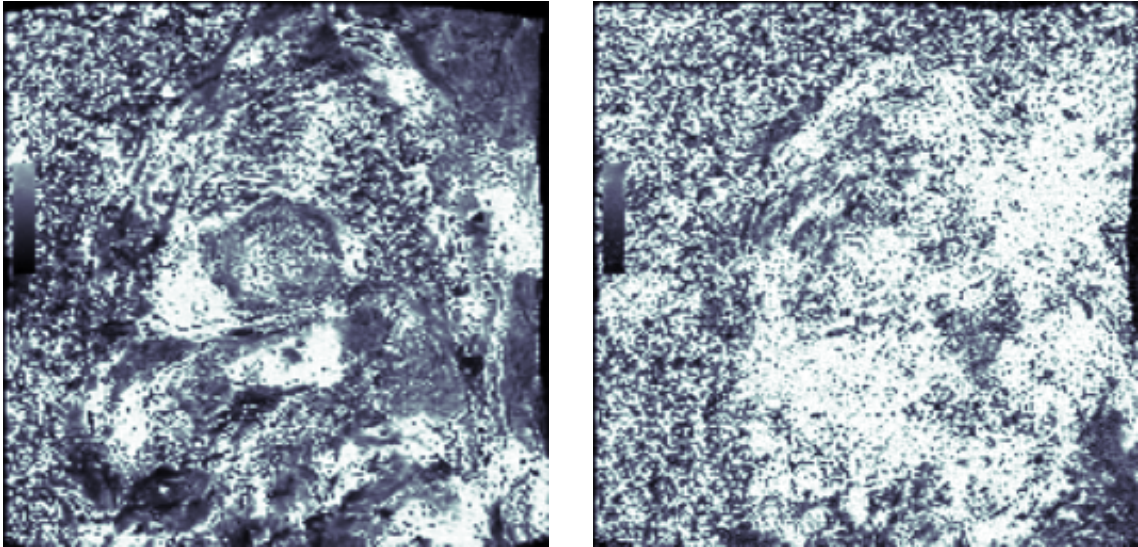
A GAN módszerhez hasonlóan itt is próbálkoztam ugyanazokkal az augmentációs technikákkal, azonban itt még rontott is a modell teljesítményén a metrikákat tekintve, FID 0.85-re nőtt. Mintavételezés során gyakran nem kivágott maszkokra is kivágott képrészleteket generált a modell.



7.6. ábra. GAN hálózattal kiértékelés során előállított, hibás, kivágott képrészlet mutató felvétel. SDM modell ilyen kiértékelése során ilyen hibát mutató képek nem fordultak elő

Azonban az augmentáció nélküli, finomhangolt modell sem generált hibátlan képeket, annak ellenére sem, hogy a metrikák alapján jelentős javulást sikerült elérni a GAN-hoz képest. A generált felvételek nagy része elég jó minőségű, viszont esetenként kifejezetten zajos képek kerülnek generálásra, amik közel felismerhetetlenek. Ezen probléma kiküszöbölése még egy későbbi feladat.

Egy másik hiányossága még a modellnek, hogy bár a generált felvételek reálisnak nem hozzáértő szemmel, az adott eredeti felvételre mégsem hasonlítanak hibátlanul legtöbb esetben. Ez abból adódik, hogy gyakran más típusú felvétel kerül



7.7. ábra. Kiértékelésnél hibás, zajos felvételek generálása

generálásra, mint az eredeti igazi MRI felvétel. Így egy továbbfejlesztési lehetősége még a modellnek, hogy vagy a maszkokba kódolva megtalálható legyen a felvétel típusa, vagy egy külön bementen lehessen azt definiálni, így explicit kontrollálva a generálási folyamatot. Természetesen ennek egy egyszerűbb megoldása lehet annyi modell tanítása, ahányféle felvétel van, de lehet, hogy azok rosszabbul teljesítenének a kisebb méretű tanító adathalmazok miatt.

8. fejezet

A megoldások értékelése

A kiértékelési metrikák alapján egyértelműen jobb felvételek generálását lehet elérni a diffúziós módszerrel, lényegesen kevesebb optimalizációval. Ennek viszont az egyik ára, hogy sokkal erősebb hardwarekkel kell rendelkezni hozzá, mivel jelentősen nagyobb modell kerül tanításra.

8.1. táblázat. A két módszer legjobb modelljeinek kiértékelése.

Módszer	FID	KID átlag	KID szórás
Pix2Pix	1.36	0.2755	0.0056
SDM	0.45	0.0383	0.0036

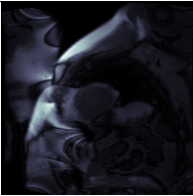
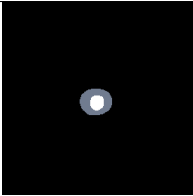
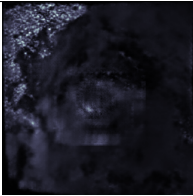
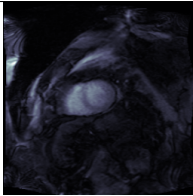
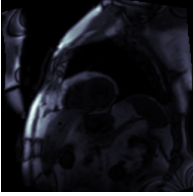

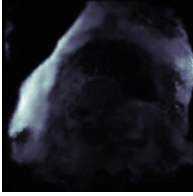
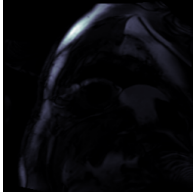
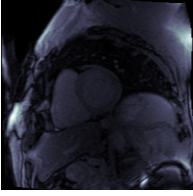
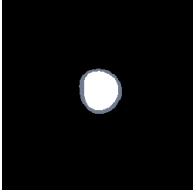
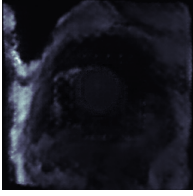
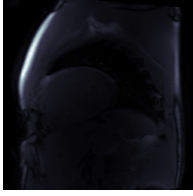
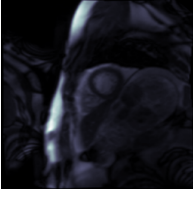

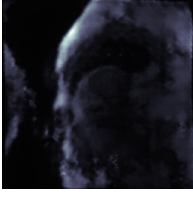
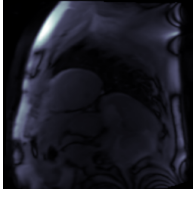
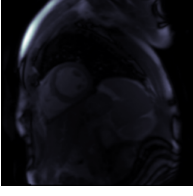

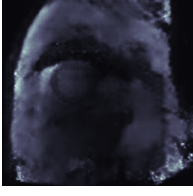
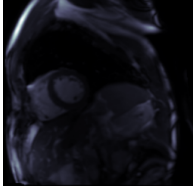
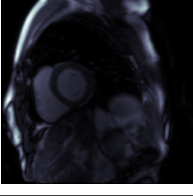

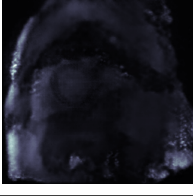
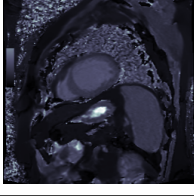
A másik hátránya szintén a modell méretéből adódik, egy-egy tanítás akár 48 óráig is eltart, míg a GAN esetén ez az idő 6 óra körül volt. Ehhez fontos hozzátenni, hogy ezek az adatok nem ugyanarról a hardwareről származtak, viszont mérések alapján a diffúziós modellhez használt Nvidia A100-as kártya többször gyorsabb, mint a GAN-nál használt Nvidia Tesla V100.

Amennyiben a feladat elvárja, hogy a mintavételezés gyorsan történjen meg, akkor a GAN-os megoldás szintén előnyösebb. SDM esetén 45 másodperc kell egy felvétel generálásához, míg Pix2Pix modell esetén ez 0.0055 másodperc, ami több mint 8000-szer gyorsabb és ezek az eredmények szintén úgy kerültek mérésre, hogy a GAN modell volt a lassabb hardwaren.

Azonban a feladatom esetében a generálás sebessége a legtöbb várható felhasználás esetében nem okozna problémát, hiszen elsősorban az adathalmazok méretének növelése lenne a cél. Ebben az alkalmazásban pedig elég egyszer legenerálni az adatokat és kimenteni, hogy később is felhasználhatóak legyenek.

Összeségében a diffúziós módszer előnyei elnyomják a hátrányait. A generált felvételek sokkal realiztikusabbnak tűnnek, nincs rajtuk az a homályos hatás, ami a GAN által létrehozott adatokon van. A tanítás folyamata is egyszerűbb, más fajta bemeneti adatokra várhatóan sokkal kevesebb optimalizációt igényelne, mint a GAN, hogy elfogadható eredményeket kapjunk.

8.2. táblázat. Felvételek összehasonlítása

Típus	Igazi felvétel	Maszk	Pix2Pix (FID: 1.36)	SDM (FID: 0.45)
T1 apex				
T1 mid				
T1 base				
T2 apex				
T2 mid				
T2 base				

9. fejezet

A módszer alkalmazási lehetőségei

Ebben a fejezetben a dolgozatom által bemutatott módszerek alkalmazási lehetőségeit vizsgálom meg. A két módszer a 4. fejezetben bemutatott rendszertervben a képgeneráló modellként kerülhetnek felhasználásra.

Ezen módszerek elsődleges alkalmazási lehetősége az adatdúsítás. Esetemben kis méretű egészségügyi adathalmazok szintetikus adatokkal való kiegészítése a cél. Azonban rendelkezik egy másodlagos alkalmazással is, anonimizált adathalmaz előállítására is alkalmasak lehetnek. Az egészségügyi adathalmazok általában nehezen hozzáférhetőek, amennyiben hozzáférhetőek akkor is előzetes anonimizálást igényelnek, hogy a páciensek személyiségi jogai ne sérüljenek a felhasználásukkal. Egy ilyen modellel, úgy lehet létrehozni nagy mennyiségű adatot és azt tanításra használni, hogy nem kell foglalkozni a páciensek miatt felmerülő problémákkal, hiszen ezek a szintetikus adatok nem köthetők egy személyhez sem.

Viszont, ahhoz hogy ezekkel a módszerekkel adathalmazokat lehessen dúsítani, fontos, hogy a modellek realiztikus felvételeket állítsanak elő kellő diverzitással. A Pix2Pix modell esetében a generált felvételekről elmondható, hogy realiztikusnak tűnnek, viszont zajosak, homályosak. Ennek ellenére is a szegmentált részek hasonlóan kivehetőek, mint az igazi felvételeken. Úgy gondolom, hogy ha egy szegmentáló modellt egy olyan adathalmazon tanítanánk, ami fel van dúsítva a GAN által generált képekkel, akkor elképzelhető, hogy jobb szegmentáló modellt tudnánk betanítani, mint csak az eredeti kis méretű adathalmazzal.

A SDM módszer szintén realiztikus felvételeket állít elő, egy nem hozzáértő ember feltehetően a legtöbbet valós felvételnek gondolná. Az adatokon nem látható az a homályos hatás, ami a GAN-nál megfigyelhető volt. Azonban ez a módszer se hibátlan mindig, előfordul, hogy zajos felvételeket generált. A "Rendszerterv" fejezetben említettem, hogy hasznos lehet minőségellenőrzést végrehajtani a szintetikus adatokon, mielőtt bekerülnek az igazi adatok mellé. Így elkerülhető lenne, hogy hibás felvételekre tanítsunk más modelleket később. Az SDM módszerrel előállított szintetikus felvételekkel feldúsított, szükség esetén szakértők bevonásával szűrt adathalmazon való betanítással, várható, hogy egy szegmentáló modell jobb teljesítményt tudna elérni.

Mindkét módszer legnagyobb gyengesége, hogy nem lehet bennük kontrollálni a felvétel típusát. A jövőben hasznos lenne, ha célzottan bizonyos típusú felvételeket lehetne generálni. Így lehetséges lenne például javítani egy imbalanced (kiegyensúlyozatlan) adathalmazt. Az itt bemutatott munka továbbfejlesztéseként ezt a funkciót, az SDM modellel érdemes megvalósítani, mivel mint a dolgozatomban

bemutattam, jobban teljesített, mint a GAN és általánosságban is elmondható a diffúziós modellekről, hogy nagyobb diverzitást tudnak lefedni egy adathalmazon belül.

10. fejezet

Összefoglalás

A dolgozatom olyan képgeneráló generatív megoldásokat mutat be, amikkel lehetséges „hamis” szív MRI felvételek generálása a bal kamra és annak falának szegmentációs maszkjából. Erre azért van szükség, mert orvosi képfelvételek általában kis mennyiségben érhetőek el, főleg esetemben szegmentációs maszkkal. Hasznos lehet egy olyan modell, ami maszkokból képes igazinak tűnő MRI felvételek generálására, ezzel feldúsítva az adathalmazokat, más feladatokhoz készülő modellek tanításához (például szegmentáló modell).

A dolgozatban először bemutattam az egészségügyi képgeneráló GAN-ok témakörét, majd hasonlóan jártam el a diffúziós modellek alkalmazásaival is. Ezek után a feladatom megoldására kiválasztottam a legígéretesebbnek tűnő megoldásokat. A GAN módszer esetén ez a Pix2Pix modell volt, míg a diffúziós modellek közül a semantic diffusion model-t alkalmaztam.

Ezek után bemutattam a felhasznált adathalmazom felépítését, a tanító és a teszt részhalmaz méretét és a különböző típusú felvételeknek az eloszlását ezekben.

Kiértékelésre mérőszámoknak a fréchet inception distance-t és a kernel inception distance-t választottam. Az én szubjektív észrevételeim mellett ezekkel mértem a modellek minőségét. Az implementációkat kiegészítettem ezekkel a mérőszámokkal, hogy objektívan ki tudjam értékelni a generált felvételeket. Ezek mellett kisebb módosításokat hajtottam még végre a projektekben, hogy felügyelni tudjam a tanításokat, illetve hogy az SDM esetén javítsak a memóriahasználaton.

Először a GAN megoldás eredményeit, tapasztalatait mutattam be. A hiperparaméterek optimalizálása mellett kísérleteztem különböző architektúrákkal és vesztességfüggvényekkel is. A tanítások végére sikerült olyan képeket generálnom, amik hasonlítanak az igazi MRI felvételekre, viszont megfigyelhető hibájuk volt, hogy a bal kamra fala gyakran zajos és a szív többi része pedig elmosódott volt.

Ezek után a diffúziós módszer tapasztalatait mutattam be. A megfelelő méretű modell és batch méret mellett ezzel a módszerrel viszonylag gyorsan lehetséges volt jobb felvételeket generálni, mint a GAN módszer esetében. Az elhomályosodás és a bal kamra zajossága már nem jelentkezett, viszont hibás adatok generálása itt is előfordult.

Végezetül összehasonlítottam a két módszer teljesítményét és alkalmazhatóságát. Összességében az SDM modellel jelentősen jobb minőségű felvételeket sikerült generálni, lényegesen kevesebb tanításból, mint a Pix2Pix modell esetén. A felvételek minősége és a könnyebb optimalizáció ellensúlyozza a hátrányait, ami elsősorban a modellhez szükséges nagy számítási teljesítményből adódik.

Feltehetőleg lehet még javítani a generált felvételek minőségein mindkét esetben, viszont az egyértelműen látszik, hogy melyik megközelítéssel érdemes hosszabb távon foglalkozni.

Köszönetnyilvánítás

Szeretném megköszönni konzulenseimnek, Kalapos Andrásnak és Dr. Gyires-Tóth Bálintnak a segítségét és iránymutatását. Ezen felül szeretnék köszönetet mondani a Városmajori Szív- és Érgyógyászati Klinikának és a Siemens Healthcare-nek, hogy biztosították a feladat elvégzéséhez szükséges adathalmazt.

Irodalomjegyzék

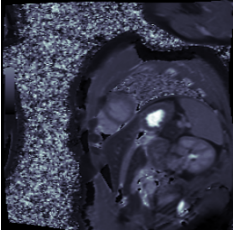

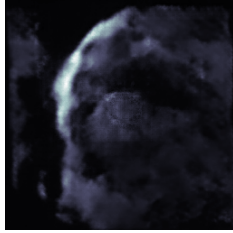
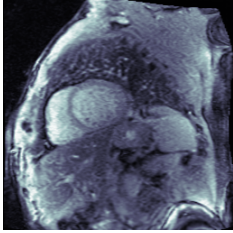
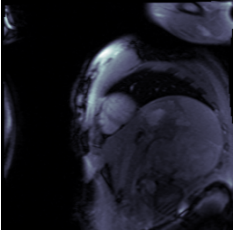

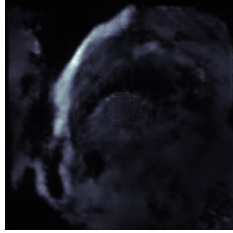
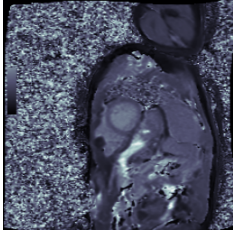
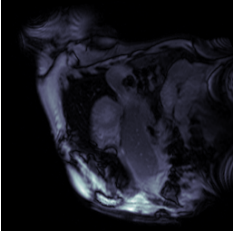
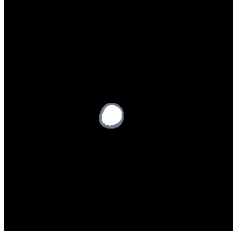
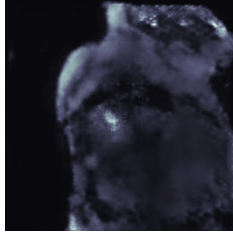
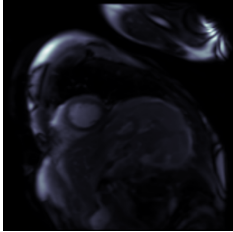
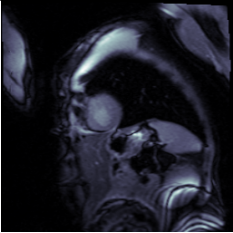
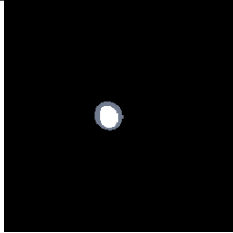
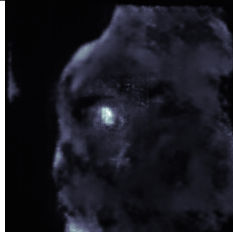
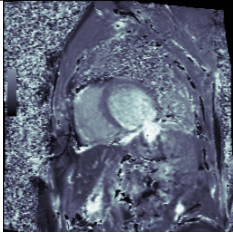
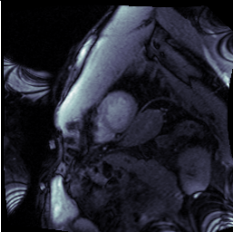
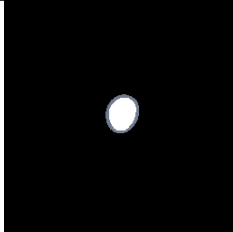
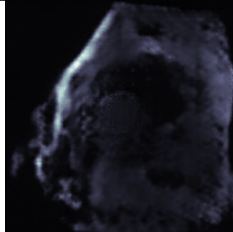
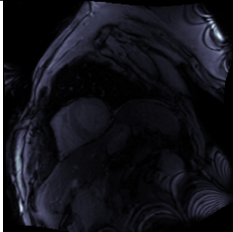
- [1] Hazrat Ali – Md Biswas – Farida Ali – Uzair Shah – Asma Alamgir – Osama Mousa – Zubair Shah: The role of generative adversarial networks in brain mri: a scoping review. *Insights into Imaging*, 13. évf. (2022. 12).
- [2] Martin Arjovsky – Soumith Chintala – Léon Bottou: Wasserstein gan, 2017.
- [3] Shane Barratt – Rishi Sharma: A note on the inception score, 2018.
- [4] Mikołaj Bińkowski – Danica J. Sutherland – Michael Arbel – Arthur Gretton: Demystifying mmd gans, 2021.
- [5] Zolnamar Dorjsembe – Sodontavilan Odonchimed – Furen Xiao: Three-dimensional medical image synthesis with denoising diffusion probabilistic models. In *Medical Imaging with Deep Learning* (konferenciaanyag). 2022.
- [6] Zolnamar Dorjsembe – Hsing-Kuo Pao – Sodontavilan Odonchimed – Furen Xiao: Conditional diffusion models for semantic 3d medical image synthesis, 2023.
- [7] Ian J. Goodfellow – Jean Pouget-Abadie – Mehdi Mirza – Bing Xu – David Warde-Farley – Sherjil Ozair – Aaron Courville – Yoshua Bengio: Generative adversarial networks, 2014.
- [8] Kaiming He – Xiangyu Zhang – Shaoqing Ren – Jian Sun: Deep residual learning for image recognition, 2015.
- [9] Martin Heusel – Hubert Ramsauer – Thomas Unterthiner – Bernhard Nessler – Sepp Hochreiter: Gans trained by a two time-scale update rule converge to a local nash equilibrium, 2018.
- [10] Phillip Isola – Jun-Yan Zhu – Tinghui Zhou – Alexei A. Efros: Image-to-image translation with conditional adversarial networks, 2018.
- [11] Tero Karras – Samuli Laine – Timo Aila: A style-based generator architecture for generative adversarial networks, 2019.
- [12] Diederik P. Kingma – Max Welling: An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12. évf. (2019) 4. sz., 307–392. p. URL [https://doi.org/10.1561%2F22000000056](https://doi.org/10.1561/2F22000000056).
- [13] Roman Macháček – Leila Mozaffari – Zahra Sepasdar – Sravanthi Parasa – Pål Halvorsen – Michael A. Riegler – Vajira Thambawita: Mask-conditioned latent diffusion for generating gastrointestinal polyp images, 2023.

- [14] Mehdi Mirza – Simon Osindero: Conditional generative adversarial nets, 2014.
- [15] Chris V. Nicholson: Generative adversarial network (gan). URL <https://wiki.pathmind.com/generative-adversarial-network-gan>.
- [16] NVIDIA Developer Blog: Improving diffusion models as an alternative to gans - part 1, 2022. URL <https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>.
- [17] Keiron O’Shea – Ryan Nash: An introduction to convolutional neural networks, 2015.
- [18] Taesung Park – Ming-Yu Liu – Ting-Chun Wang – Jun-Yan Zhu: Semantic image synthesis with spatially-adaptive normalization, 2019.
- [19] Wei Peng – Ehsan Adeli – Tomas Bosschieter – Sang Hyun Park – Qingyu Zhao – Kilian M. Pohl: Generating realistic brain mris via a conditional diffusion probabilistic model, 2023.
- [20] Minh Phan – Zhibin Liao – Johan Verjans – Minh-Son To: Structure-preserving synthesis: Maskgan for unpaired mr-ct translation, 2023. 07.
- [21] Olaf Ronneberger – Philipp Fischer – Thomas Brox: U-net: Convolutional networks for biomedical image segmentation, 2015.
- [22] Hoo-Chang Shin – Neil A Tenenholtz – Jameson K Rogers – Christopher G Schwarz – Matthew L Senjem – Jeffrey L Gunter – Katherine Andriole – Mark Michalski: Medical image synthesis for data augmentation and anonymization using generative adversarial networks, 2018.
- [23] Youssef Skandarani – Alain Lalande – Jonathan Afilalo – Pierre-Marc Jodoin: Generative adversarial networks in cardiology. *Canadian Journal of Cardiology*, 38. évf. (2022) 2. sz., 196–203. p. ISSN 0828-282X. URL <https://www.sciencedirect.com/science/article/pii/S0828282X21008606>. Focus Issue: New Digital Technologies in Cardiology.
- [24] Youssef Skandarani – Nathan Painchaud – Pierre-Marc Jodoin – Alain Lalande: On the effectiveness of gan generated cardiac mris for segmentation, 2020.
- [25] Jascha Sohl-Dickstein – Eric A. Weiss – Niru Maheswaranathan – Surya Ganguli: Deep unsupervised learning using nonequilibrium thermodynamics, 2015.
- [26] David Stojanovski – Uxio Hermida – Pablo Lamata – Arian Beqiri – Alberto Gomez: Echo from noise: synthetic ultrasound image generation using diffusion models for real image segmentation, 2023.
- [27] Weilun Wang – Jianmin Bao – Wengang Zhou – Dongdong Chen – Dong Chen – Lu Yuan – Houqiang Li: Semantic image synthesis via diffusion models, 2022.
- [28] Jun-Yan Zhu – Taesung Park – Phillip Isola – Alexei A. Efros: Unpaired image-to-image translation using cycle-consistent adversarial networks, 2020.

Függelék

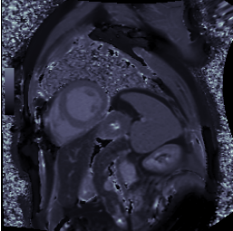
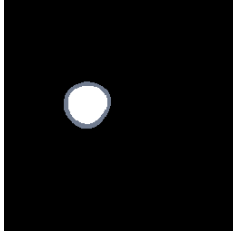
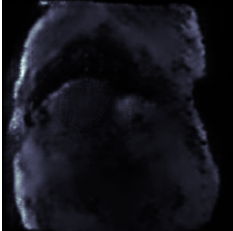
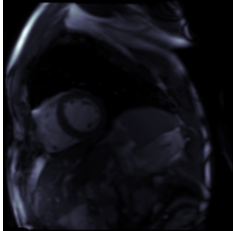
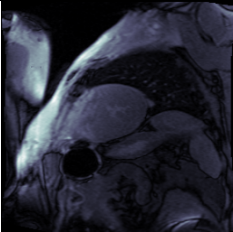
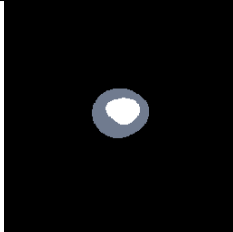
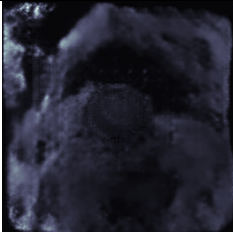
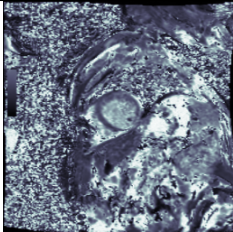
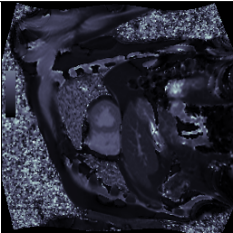
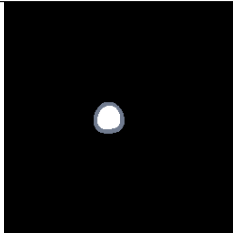
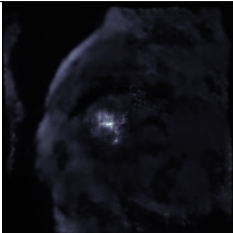
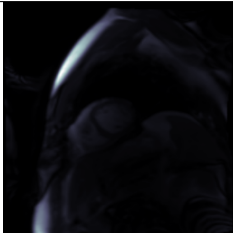
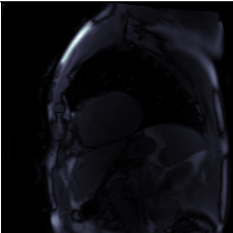
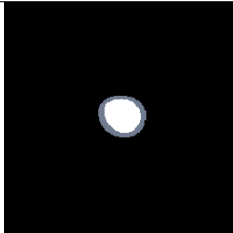
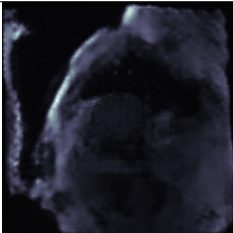
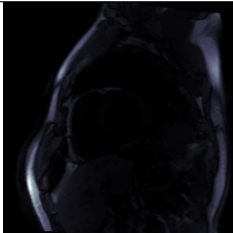
F.1. T1 apex

F.1. táblázat. T1 apex felvételek összehasonlítása

Igazi felvétel	Maszk	Pix2Pix (FID: 1.36)	SDM (FID: 0.45)
			
			
			
			
			

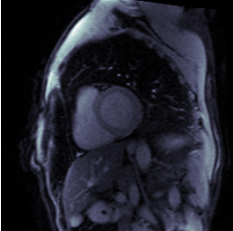

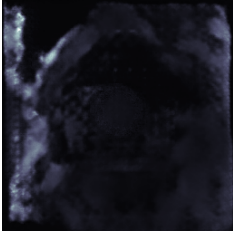
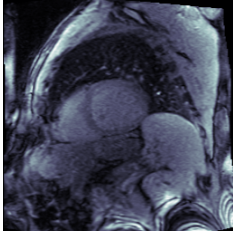
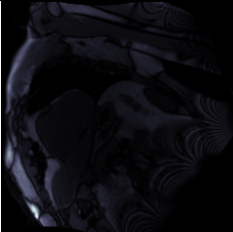

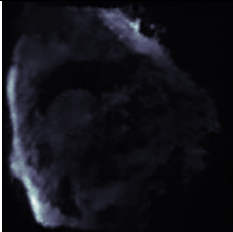
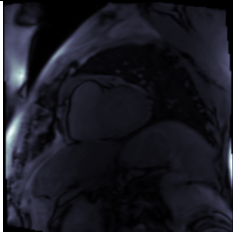
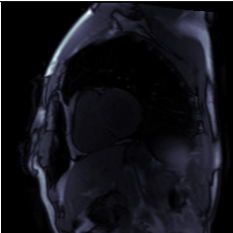
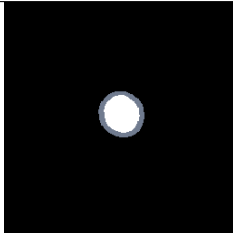
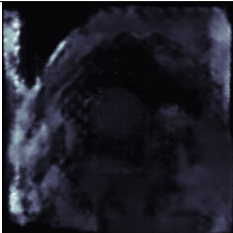
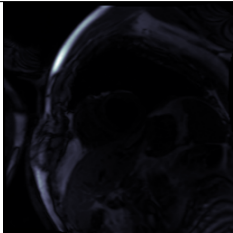
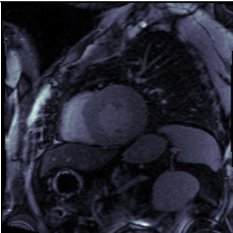
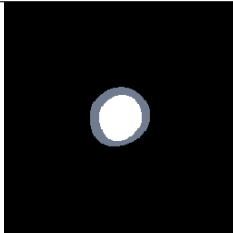
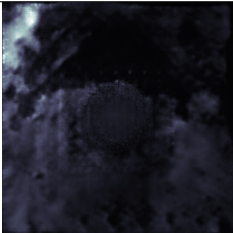
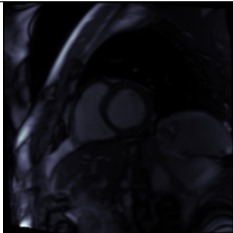
F.2. T1 mid

F.2. táblázat. T1 mid felvételek összehasonlítása

Igazi felvétel	Maszk	Pix2Pix (FID: 1.36)	SDM (FID: 0.45)
			
			
			
			

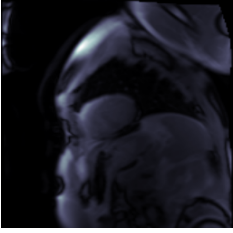
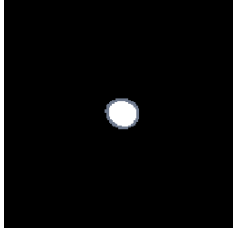
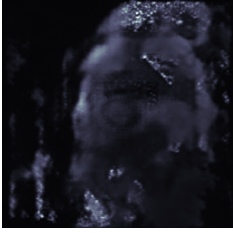
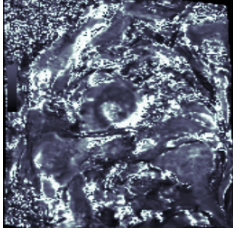
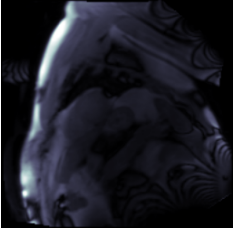
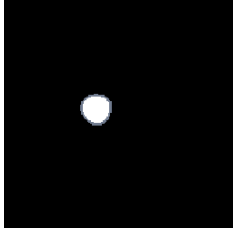
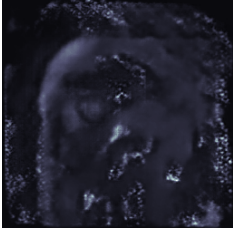
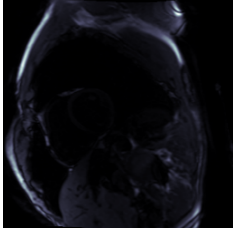
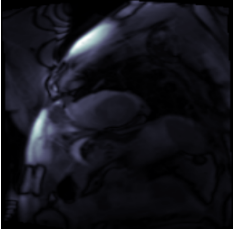
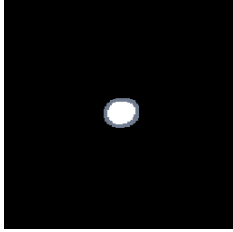
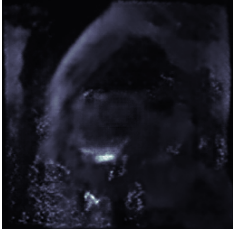
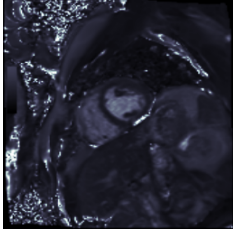
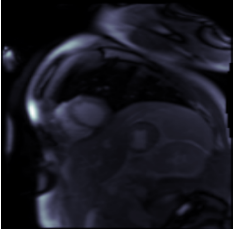

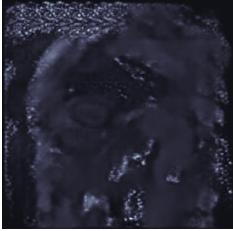
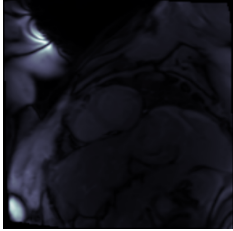
F.3. T1 base

F.3. táblázat. T1 base felvételek összehasonlítása

Igazi felvétel	Maszk	Pix2Pix (FID: 1.36)	SDM (FID: 0.45)
			
			
			
			

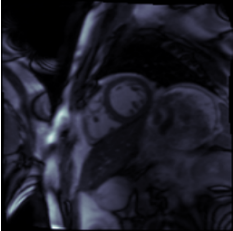
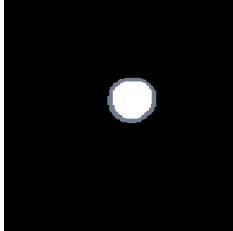

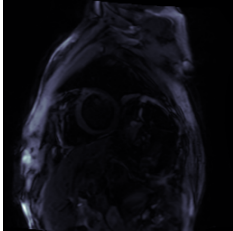
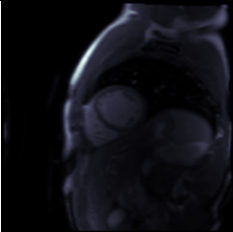
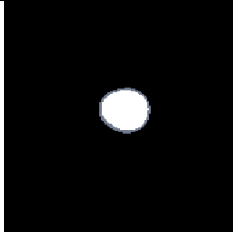
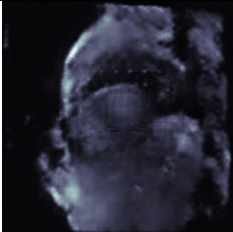
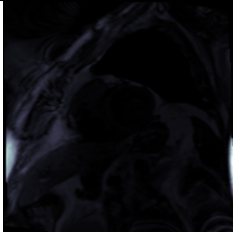
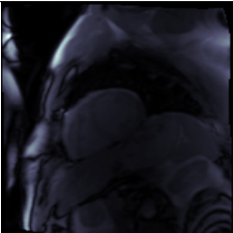
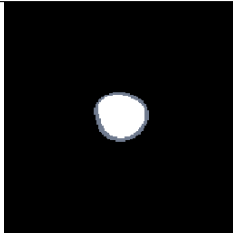
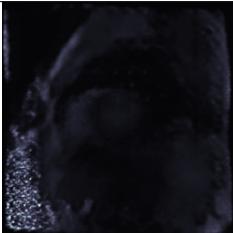
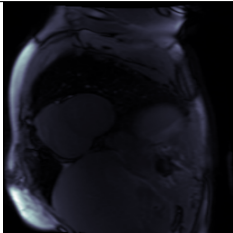
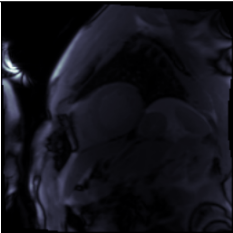

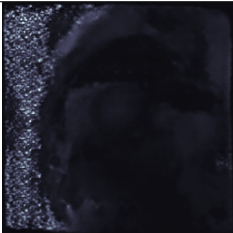
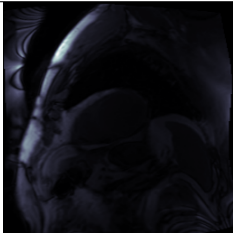
F.4. T2 apex

F.4. táblázat. T2 apex felvételek összehasonlítása

Igazi felvétel	Maszk	Pix2Pix (FID: 1.36)	SDM (FID: 0.45)
			
			
			
			

F.5. T2 mid

F.5. táblázat. T2 mid felvételek összehasonlítása

Igazi felvétel	Maszk	Pix2Pix (FID: 1.36)	SDM (FID: 0.45)
			
			
			
			

F.6. T2 base

F.6. táblázat. T2 base felvételek összehasonlítása

Igazi felvétel	Maszk	Pix2Pix (FID: 1.36)	SDM (FID: 0.45)
