



Universiteit  
Leiden

The Netherlands

## Deep learning for automated analysis of cardiac imaging: applications in Cine and 4D flow MRI

Sun, X.

### Citation

Sun, X. (2023, July 5). *Deep learning for automated analysis of cardiac imaging: applications in Cine and 4D flow MRI*. Retrieved from <https://hdl.handle.net/1887/3629578>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3629578>

**Note:** To cite this publication please use the final published version (if applicable).

## Samenvatting en toekomstig werk

Cine en 4D flow cardiale MRI zijn twee belangrijke niet-invasieve MR-beeldvormingstechnieken om de hartfunctie te beoordelen en cardiovasculaire ziekten te diagnosticeren. Cine MRI biedt grote details van het zachte weefsel, waardoor klinische deskundigen de structuur en functie van het hart kunnen beoordelen. 4D flow MRI heeft verder de mogelijkheid tot het maken van driedimensionale tijdsopnamen van de bloedstroomsnelheid, die kan worden gebruikt om intra-cardiale hemodynamische parameters af te leiden. In dit proefschrift hebben we op deep learning gebaseerde benaderingen ontwikkeld voor het analyseren van cine en 4D flow cardiale MRI. In dit hoofdstuk vatten we de voorgaande hoofdstukken samen en bespreken we mogelijke richtingen voor toekomstig werk.

### Samenvatting

In **hoofdstuk 1** hebben we een algemene inleiding gegeven over cine en 4D flow cardiale MRI en deep learning toepassingen op het gebied van cardiale MRI. In hoofdstuk 2 hebben we een netwerk voor bemonsteringsinspectie voorgesteld dat speciaal ontworpen methoden voor datavergroting combineert om de CMR-beeldkwaliteit te beoordelen. De voorgestelde methode presteerde concurrerend ten opzichte van de andere methoden in de CMRxMotion challenge. In hoofdstuk 3 hebben we temporele en ruimtelijke stacks voorgesteld om temporele of ruimtelijke informatie op te nemen met behulp van het stack-aandachtsmechanisme voor segmentatie van de linkerventrikel in korte-as cine MRI beelden. In hoofdstuk 4 hebben we het concept van domein generalisatie verder bestudeerd in de setting van rechterventrikel segmentatie in ongeziene datasets, zoals data met verschillen in acquisitie protocol, over verschillende centra, scanner leveranciers en ziekte beelden. In hoofdstuk 5 onderzochten wij de haalbaarheid van op deep learning gebaseerde benaderingen om de LV rechtstreeks te segmenteren op basis van 4D flow-MRI en onderzochten wij de prestaties van de integratie van kenmerken die werden geëxtraheerd uit magnitude- en snelheidsbeelden. In hoofdstuk 6 werd een op *Transformers* gebaseerd feature fusie model ontwikkeld om de prestaties van LV-segmentatie van 4D flow MRI te verbeteren. In hoofdstuk 7 werd een CNN-model getraind om de bloedstroomsnelheid te voorspellen op basis van lange-as cine-MRI beelden, waarbij de corresponderende 4D-flow gegevens als referentie werden gebruikt.

**Hoofdstuk 2** CMR kan last hebben van bewegingsgerelateerde artefacten die resulteren in beelden van niet-diagnostische kwaliteit. Visuele inspectie van de beeldkwaliteit is tijdrovend en bovendien afhankelijk van ervaren radiologen. In dit

hoofdstuk stellen we een automatische methode voor om de kwaliteit van CMR-beelden te beoordelen. Met beperkte hoeveelheid data en een onevenwichtige klassenverhouding hebben wij drie speciaal ontworpen datavergrotingsmethoden voorgesteld om de dataset uit te breiden, waaronder het genereren van overgangsfasen tussen de ED- en ES-fasen, het genereren van beelden met verschillende mate van ademhalingsbeweging en het genereren van beelden met behulp van histogrammatching en lineaire interpolatie. Om de bemonsteringsinspectie na te bootsen, nemen we willekeurig twee deelmonsters van één 3D-volume om de kwaliteit van een 3D-volume te schatten. In het ontwikkelde model, dat werd aangepast aan ResNet, wordt kanaalaandacht gebruikt om de intrakanaalrelatie te onderzoeken voor de kenmerken die uit elk deelmonster worden geëxtraheerd. Vervolgens wordt een feature fusie module geïntroduceerd om features van twee subsamples te fuseren om de beeldkwaliteit te voorspellen. De voorgestelde methode is gevalideerd in de CMRxMotion-wedstrijd van 2022 en behaalde een gemiddelde nauwkeurigheid van 75% en 72,5% in respectievelijk de training- en validatiedataset. Bovendien eindigde onze methode op de vierde plaats in de testdataset die door de organisator verborgen was gehouden.

**Hoofdstuk 3** In dit hoofdstuk gebruiken we de spatio-temporele informatie van naburige beelden om de nauwkeurigheid van segmentatie te verbeteren. Het doelbeeld wordt gestapeld met zijn ruimtelijke of temporele naburige beelden als input. Vervolgens wordt een stapel aandacht ontwikkeld om de relevante kenmerken te extraheren en te wegen met het doelbeeld als leidraad. De stackaandacht wordt in U-Net ingevoegd om automatisch de LV en het myocard te segmenteren uit multi-slice korte-as cardiale MRI. Een interne dataset van één centrum en een openbare dataset van de 2017 Automated Cardiac Segmentation Challenge (ACDC) werden betrokken bij het evalueren en valideren van de voorgestelde methode. Het model is eerst getraind op de interne dataset en vervolgens verfijnd op de openbare dataset. De methode behaalde een Dice van 0,91 en een Hausdorff Distance van 3,37 mm op de interne dataset. De prestaties op de ACDC-dataset bereikten een Dice van 0,92 en 0,89 en een Hausdorff Distance van 9,7 mm en 7,1 mm voor respectievelijk de ED- en ES-fasen, wat een goede generalisatie bevestigt. Bovendien laten de resultaten in beide datasets een hoge correlatie zien van LVEF en myocardmassa afgeleid van het model en handmatige segmentatie, wat een potentieel waardevolle toepassing in de klinische praktijk aantoont.

**Hoofdstuk 4** Dit hoofdstuk richt zich op modelgeneralisatie, waarbij het doel is een model te ontwikkelen dat goed presteert op ongeziene datasets van verschillende centra, leveranciers of verschillende ziekten. De *M&Ms-2-challenge* is gemotiveerd om de rechterhartkamer te segmenteren op basis van een steekproef van 360 cardiale MRI-datasets met meerdere ziekten, meerdere beeld oriëntaties en meerdere centra. De meest eenvoudige aanpak van dit probleem is het verzamelen van meer data om

een model te trainen. Met beperkte gelabelde data introduceren we eerst een op intensiteit gebaseerde registratiemethode om de beschikbare labels van de eind-diastolische (ED) en eind-systolische (ES) fasen te propageren naar de andere ongelabelde fasen. Vervolgens onderzoeken wij de prestaties van verschillende invoermodaliteiten, waaronder een enkel 2D-beeld, een meerkanaals 2D-beeld en een 3D-volume. Het meerkanaals 2D-beeld wordt opgebouwd met behulp van de in hoofdstuk 3 voorgestelde ruimtelijke en temporele stapeling. Op de validatiedataset behaalde onze methode een Dice van 0,92 en 0,92, Hausdorff Distance van 9,5 mm en 5,3 mm in respectievelijk de korte en lange as beelden. Onze methode levert ook goede prestaties op de verborgen testdataset, met een Dice van 0,93 en 0,92 en een Hausdorff Distance van 10,6 mm en 6,0 mm in respectievelijk korte- en lange-as aanzicht. De experimentele resultaten tonen aan dat het meerkanaals 2D-beeld meer informatie biedt voor de segmentatie. De combinatie van volume-invoer en labelpropagatie kan het generalisatievermogen verder verbeteren.

Eerder gerapporteerde benaderingen voor 4D-flow segmentatie zijn gebaseerd op de registratie tussen cine-MRI en 4D-flow gegevens, wat hoge rekencapaciteit vereist. Hoofdstuk 5 en Hoofdstuk 6 richten zich op LV-segmentatie direct vanuit 4D flow MRI zonder afhankelijk te zijn van aanvullende cine MRI. In **Hoofdstuk 5**, hebben we onderzocht met behulp van de combinatie van magnitude- en snelheidsbeelden samen met 4D-flow gegevens als invoer. Het slechte contrast tussen de hartkamers en het myocard zal resulteren in inherente onzekerheid in de segmentatieresultaten. Daarom wordt de Monte Carlo-uitvalmethode geïntroduceerd om de segmentatie-onzekerheid te beoordelen. Daarnaast worden vijf op deep learning gebaseerde modellen vergeleken om het effect te onderzoeken van het gebruik van verschillende netwerkarchitecturen, datavorverwerking, invoer en functiefusiemethoden op de segmentatieprestaties. Op basis van de resultaten bleek de voorgestelde methode zeer nauwkeurig te zijn. Bovendien vertonen de klinische parameters die zijn afgeleid van het beste model een hoge correlatie met resultaten die zijn afgeleid van handmatige annotaties, wat de haalbaarheid van LV-segmentatie rechtstreeks uit 4D flow MRI bevestigt.

**Hoofdstuk 6** presenteert een op een *Transformer* gebaseerde efficiënte kenmerkfusiemethode om de informatie uit magnitude- en snelheidsbeelden te fuseren en om de segmentatieprestaties in 4D flow MRI te verbeteren. Het netwerk is een encoder-decoderstructuur gebaseerd op U-Net. In de encoder worden de magnitude- en snelheidsbeelden beschouwd als de invoer van twee afzonderlijke takken. De functies van hetzelfde niveau zijn geïntegreerd met behulp van de feature fusie-module. De *cross-* en *self-fusion*-laag in de feature fusie-module is bedoeld om de inter- en intra-relatie tussen die features te verkennen. De gefuseerde kenmerken worden toegevoegd aan de originele kenmerken. De gepaarde kenmerken op meerdere niveaus worden aaneengeschaakeld langs de kanaaldimensie, gevolgd door

een convolutionele laag als invoer van de decoder. De decoder wordt hetzelfde gehouden als die in U-Net. De voorgestelde methoden leveren de beste prestaties in vergelijking met de andere modellen en zorgen voor een significante verbetering van de klinische parameters, resulterend in een Pearson-correlatiecoëfficiënt van respectievelijk 83,3%, 97,4%, 96,97% en 98,92% voor LVEF, EDV, ESV en KE. De voorgestelde feature fusie-methode maakt het daarom mogelijk om de features van verschillende modaliteiten op een efficiënte manier samen te voegen.

**Hoofdstuk 7** In dit hoofdstuk hebben we een op deep learning gebaseerde methode ontworpen en geëvalueerd om het intra-cardiale bloedstroompatroon te voorspellen op basis van cine-MRI lange as opnamen, waarbij we de snelheden afgeleid van 4D-flow gegevens gebruiken als referentie. Het netwerk, een variant van U-Net en ResNet, neemt een reeks cine-MR-beelden als input om de verplaatsing van bloed over de cardiale frames te extraheren. Hoewel werd aangetoond dat de gemiddelde voorspelde snelheid met 26,69% werd onderschat, vertoont het globale in de tijd variërende bloedstroompatroon een hoge correlatie met de van de 4D-flow afgeleide snelheden. Een mogelijke toepassing van de voorgestelde methode is het schatten van de E/A-ratio. De resultaten gaven aan dat de E/A-ratio zonder significante bias kan worden geschat en daarnaast de diastolische functie kan classificeren met een hoge nauwkeurigheid. Onze studie is de eerste die deep learning gebruikt voor de voorspelling van de bloedstroom op basis van cine-MRI. Na verdere verbetering van het model zou dit werk potentieel waardevol kunnen zijn in klinische toepassingen om de intracardiale bloedstroom te visualiseren zonder aanvullende 4D-flow gegevens.

## **Discussie en toekomstig werk**

Het werk dat in dit proefschrift wordt gepresenteerd, heeft tot doel op deep learning gebaseerde methoden te ontwikkelen voor geautomatiseerde analyse van cardiale cine en 4D flow MRI.

De in hoofdstuk 3 en 4 ontwikkelde netwerken waren voornamelijk gericht op geautomatiseerde segmentatie in cine MRI. In hoofdstuk 3 hebben wij aangetoond dat het extraheren van temporele of ruimtelijke informatie uit naburige coupes de segmentatieprestatie in korte-as cine MRI kan verbeteren. De prestaties van het gebruik van ruimtelijke kenmerken zijn beter dan die van temporele informatie. De ruimtelijke stack kan namelijk meer informatie verschaffen over de positie, grootte en vorm van het hart. Terwijl de beelden in de temporele stack op elkaar lijken en vergelijkbare kenmerken bevatten. In deze studies heeft deep learning zijn veelbelovende toepassingen in cardiale MRI-segmentatie aangetoond. De ontwikkelde benaderingen zijn echter gevalideerd in gevallen waarin de testdata afkomstig zijn uit hetzelfde domein als de trainingsdata. In een realistisch scenario

kan een aanzienlijke prestatiedaling worden waargenomen wanneer een getraind model wordt toegepast op data uit een ander domein. Wanneer ons op de Leeds University dataset (LUD) getrainde model bijvoorbeeld rechtstreeks wordt toegepast op de ACDC dataset, daalt de segmentatienauwkeurigheid van 90% naar 70%. Dit kan worden verklaard door de populatie verschillen van verschillende locaties, leeftijden, geslachten, rassen en pathologieën, en verschillen in beeldvorming door verschillende scanner leveranciers, protocollen en magnetische veldsterktes die resulteren in heterogeniteit van de dataverdeling. De heterogeniteit kan niet volledig worden geëlimineerd door voorbewerking van de data. Het model moet worden verfijnd of opnieuw worden getraind op de nieuwe datareeks om goede prestaties te bereiken. Daarom is domeingeneralisatie een technisch knelpunt voor de toepassing van deep learning in realistische klinische omgeving. Het verzamelen en labelen van een grote hoeveelheid data van verschillende centra en leveranciers is de meest eenvoudige oplossing. Het is echter kostentechnisch onhaalbaar om voor elk domein handmatige annotaties van hoge kwaliteit te verkrijgen, omdat daarvoor expert kennis nodig is, en het is ook onmogelijk om het volledige spectrum van data te bestrijken. In hoofdstuk 4 introduceerden wij registratie om de beschikbare segmentatielabels door te geven aan ongelabelde beelden om de hoeveelheid trainingsdata te vergroten. Bovendien worden technieken voor datauitbreiding gebruikt om de verscheidenheid aan trainingsdata te vergroten om de robuustheid van het model te verbeteren. De resultaten tonen aan dat het model goed generaliseert op data met ongeziene pathologieën. Als veelbelovende richting voor toekomstig onderzoek is het ook de moeite waard *selflearning* en semi-gesuperviseerd leren te onderzoeken om meer voorkennis te extraheren om het generalisatievermogen van het model te verbeteren wanneer de hoeveelheid trainingsdata beperkt is.

In hoofdstuk 5 vergelijken we verschillende modellen voor LV-segmentatie rechtstreeks uit 4D flow-MRI zonder gebruik te maken van de registratie tussen cine en 4D flow-MRI. In hoofdstuk 6 verbeteren we de prestaties verder met behulp van een nieuwe feature fusie methode. Er zijn echter ook andere overwegingen waarmee rekening moet worden gehouden. Ten eerste, hoewel de meeste studies zich richten op de invoering van nieuwe algoritmen, is de voorbewerking van de data, waaronder correctie, verbetering, resampling en normalisatie, ook van groot belang. Zo blijkt uit hoofdstuk 5 dat de prestaties van herschikte datavolumes beter zijn dan die van de oorspronkelijke data zonder datavoorbewerking. Evenzo overtreft het model van nnUnet (*no new Unet*), een zelfconfigurerende methode, de meeste bestaande op deep learning gebaseerde segmentatiebenaderingen op 23 openbare datasets. De goede prestaties worden niet bereikt door de invoering van een nieuwe netwerkarchitectuur voor elk type data, maar zijn het resultaat van het zorgvuldig ontworpen proces van automatische zelfconfiguratie. Ten tweede moeten de uiteindelijke evaluatiemetingen waardevol en betrouwbaar zijn voor de

kwantitatieve beoordeling van de prestaties van het model. De in hoofdstuk 5 voorgestelde modellen behaalden vergelijkbare resultaten voor wat betreft Dice en ASD, zoals gerapporteerd in tabel 5.2, wat het moeilijk maakt het beste model te selecteren. In het algemeen leveren de klinisch relevante metrieken, afgeleid van de segmentatieresultaten, zinvolle en bruikbare informatie op voor diagnose en behandeling. In vergelijking met de andere state-of-arts verbetert de voorgestelde methode in hoofdstuk 6 de Dice met slechts 2%, maar de Pearson correlatiecoëfficiënt in EDV, ESV en KE werd verbeterd met 9%, 7% en 16%. Daarom moeten de klinische parameters worden betrokken bij de vergelijking van de prestaties van verschillende modellen om ervoor te zorgen dat de algoritmen betrouwbaar zijn voor gebruik in medische toepassingen. De mogelijkheid bestaat om op deep learning gebaseerde *multi-task* netwerken te ontwikkelen om gezamenlijk de taak van hartsegmentatie en de regressie van volume of ejectiefractievoorspelling uit te voeren. Ten derde werd in hoofdstuk 6 een op *Transformers* gebaseerde feature fusie module gepresenteerd die de beste prestaties leverde. De module kan informatie uit twee verschillende modaliteiten of beeld oriëntaties efficiënt integreren. De module kan worden aangepast voor andere toepassingen, zoals de integratie van korte-as en lange-as cine-MRI voor automatische diagnostoepassing, of de combinatie van apicale vier-kamer en twee-kamer opnames in echocardiografische beelddata.

## **Algemene conclusies**

Concluderend stelt dit proefschrift deep learning-gebaseerde methoden voor om cardiale MRI te kwantificeren. De beschreven methoden kunnen worden toegepast voor cine MR-beeldkwaliteitsclassificatie en ventrikelsegmentatie zonder humane interactie. Onderzoek naar het combineren en samenvoegen van magnitude- en snelheidsbeelden kan nuttig zijn voor de segmentatie van de linkerventrikel in 4D flow-MRI, wat nog niet volledig is onderzocht. Bovendien hebben wij een netwerk voorgesteld om het bloedstroompatroon te voorspellen op basis van de cine MRI. Door de combinatie van visualisatie van de bloedstroom en myocardiale beweging in een routinematig verkregen standaard CMR-onderzoek, kan de methode mogelijk worden gebruikt in klinische studies. Alle in dit proefschrift beschreven deep learning-methoden werden geëvalueerd op MRI-beelddata, maar kunnen potentieel ook worden toegepast op andere beeldvormingsmodaliteiten zoals computertomografie en echocardiografie.