



Universiteit  
Leiden  
The Netherlands

## Supervised learning in medical image registration

Sokooti, H.

### Citation

Sokooti, H. (2021, November 22). *Supervised learning in medical image registration*. *ASCI dissertation series*. Retrieved from <https://hdl.handle.net/1887/3243762>

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/1887/3243762>

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

## Samenvatting en toekomstig werk

Beeldregistratie is een cruciale taak in de medische beeldverwerking. Het uitvoeren van een snelle automatische beeldregistratie met minder handmatige fijnafstelling kan tal van medische beeldverwerkingsprocedures versnellen. Bovendien kan een automatische kwaliteitsbeoordeling van de registratie deze tijdrovende taak versnellen. In dit proefschrift hebben we een op leren gebaseerde beeldregistratietechniek ontwikkeld, genaamd RegNet. Bovendien hebben we twee kwaliteitsbeoordelingsmechanismen voorgesteld, gebruikmakend van random forests (RF) en convolutional long short term memory (ConvLSTM), waarbij de laatste sneller en nauwkeuriger presteert. In dit hoofdstuk vatten we de vorige hoofdstukken samen en bespreken we mogelijke richtingen voor vervolgonderzoek.

### Samenvatting

In het eerste hoofdstuk hebben we algemene informatie gegeven over beeldregistratie en kwaliteitsbeoordeling van registratie. In Hoofdstuk 2 stellen we een convolutional neural network architectuur voor om niet-rigide beeldregistratie op te lossen door middel van een leerbenadering. Het voorgestelde RegNet wordt getraind met behulp van een reeks willekeurig kunstmatig gegenereerde DVF's met een maximale vervorming van 8 mm in elke richting. In Hoofdstuk 3 hebben we het voorgestelde RegNet substantieel verbeterd door een meerfasige benadering te gebruiken en de kunstmatige procedure voor het genereren van de DVF's te verbeteren. Een kwantitatieve foutvoorspelling van medische beeldregistratie wordt voorgesteld in Hoofdstuk 4 met behulp van een regressie forest. Deze random forest is opgebouwd met kenmerken die betrekking hebben op het transformatiemodel en kenmerken die verband houden met de ongelijkheid na registratie op onderscheidende landmark locaties. In Hoofdstuk 5 wordt een hiërarchische voorspelling van de registratiefout voorgesteld met behulp van een convolutional LSTM met toepassing op CT scans van de borstkas. De voorgestelde methode is aanzienlijk sneller dan methoden die meerdere registraties nodig hebben.

**Hoofdstuk 2** In dit hoofdstuk stellen we een methode voor om niet-rigide beeldregistratie op te lossen door middel van een leerbenadering, in plaats van via iteratieve optimalisatie van een vooraf gedefinieerde ongelijkheidsmetriek. We ontwerpen

een Convolutional Neural Network (CNN)-architectuur die, in tegenstelling tot al het andere werk, het verplaatsingsvectorveld (DVF) rechtstreeks schat op basis van een paar invoerbeelden. Dit hoofdstuk is een van de eerste voorgestelde methoden voor niet-rigide DL-gebaseerde beeldregistratie. Het voorgestelde RegNet wordt getraind met behulp van een reeks willekeurig kunstmatig gegenereerde DVF's met een maximale vervorming van 8 mm in elke richting. De voorgestelde methode definieert niet expliciet een ongelijkheidsmetriek en integreert beeldinhoud op meerdere schalen om het network uit te rusten met contextuele informatie. Tijdens het testen wordt niet-rigide registratie in één keer uitgevoerd, in tegenstelling tot de huidige iteratieve methoden. We hebben RegNet getest met een interne dataset van CT scans van de borstkas, SPREAD genaamd. RegNet behaalde een gemiddelde doelregistratiefout (Target Registration Error, TRE) van 1,66 mm over de testcases. De resultaten laten zien dat de nauwkeurigheid van RegNet vergelijkbaar is met een conventionele B-spline-registratie, voor anatomie binnen het bereik, d.w.z. minder dan 8 mm.

In **Hoofdstuk 3** verbeteren we de oorspronkelijke RegNet-methode die in Hoofdstuk 2 is geïntroduceerd aanzienlijk. De nieuw voorgestelde methode maakt gebruik van een meertrapsbenadering, waardoor het bereik aanzienlijk wordt vergroot. De kunstmatige generatie van de DVF's is verbeterd door meer algemene vervormingen op te nemen, evenals meer realistische vervormingen zoals ademhalingsbewegingen. We hebben geëxperimenteerd met verschillende network architecturen en het voorgestelde "U-Net-advanced"-ontwerp leverde betere prestaties in de validatieset. Dit ontwerp was vergelijkbaar met een U-Net, maar met toevoeging van gedilateerde convolutional lagen. De voorgestelde methode, RegNet, wordt geëvalueerd op meerdere databases van CT-scans van de borstkas en behaalde een doelregistratiefout van  $2,32 \pm 5,33$  mm en  $1,86 \pm 2,12$  mm op SPREAD en DIR-Lab- 4DCT-onderzoeken, respectievelijk. Bijgevolg behaalde het verbeterde RegNet het beste resultaat in de DIR-Lab 4DCT-studie van alle gepubliceerde op DL-gebaseerde registratiemethoden. De gemiddelde inferentietijd van RegNet met twee fasen is ongeveer 2,2 s.

**Hoofdstuk 4** presenteert een kwantitatieve foutvoorspelling van medische beeldregistratie met behulp van een regressie forest. Een nieuwe automatische methode wordt voorgesteld om de registratiefout op een kwantitatieve manier te voorspellen, en wordt toegepast op CT-scans van de borstkas. Om de registratiefout lokaal te voorspellen, wordt gebruik gemaakt van een random regressie forest. Het forest is gebouwd met kenmerken die betrekking hebben op het transformatiemodel en kenmerken die verband houden met de ongelijkheid na registratie. De kenmerkenset bestaat uit de variatie van het verplaatsingsvectorveld, de variatiecoëfficiënt van gezamenlijke histogrammen, determinant van de Jacobiaan, de modality independent neighborhood descriptor (MIND) en de lokale genormaliseerde wederzijdse informatie. Het forest wordt getraind en getest met behulp van handmatig geannoteerde corresponderende

punten tussen paren CT-scans van de borstkas in twee experimenten: SPREAD (getraind en getest op SPREAD) en inter-database (met drie databases SPREAD, DIR-Lab-4DCT en DIR-Lab-COPDgene). De resultaten laten zien dat de gemiddelde absolute regressiefouten  $1,07 \pm 1,86$  en  $1,76 \pm 2,59$  mm zijn voor respectievelijk het SPREAD- en het interdatabase-experiment. De algehele nauwkeurigheid van classificatie in drie klassen (correct, slecht en verkeerde registratie) is 90,7% en 75,4%, voor respectievelijk SPREAD en inter-database. De goede prestaties van de voorgestelde methode maken belangrijke toepassingen mogelijk, zoals automatische kwaliteitscontrole bij grootschalige beeldanalyse.

In **Hoofdstuk 5** wordt een hiërarchische voorspelling van verkeerde uitlijning van de registratie voorgesteld met behulp van een convolutional LSTM met toepassing op CT-scans van de borstkas. De voorgestelde methode is aanzienlijk sneller dan methodes met meerdere registraties. Deze taak wordt geprojecteerd naar een classificatieprobleem met meerdere klassen van verkeerde uitlijning: “correct” 0-3 mm, “slecht” 3-6 mm en “verkeerd” meer dan 6 mm. In plaats van een directe voorspelling stellen we een hiërarchische benadering voor, waarbij de voorspelling geleidelijk wordt verfijnd van grof naar fijn. Onze oplossing is gebaseerd op een convolutional Long Short Term Memory (LSTM), dat hiërarchische voorspellingen van verkeerde uitlijning gebruikt op drie resoluties van het beeldpaar, waarbij gebruik wordt gemaakt van de intrinsieke sterke punten van een LSTM voor dit probleem. De convolutional LSTM wordt getraind op een set kunstmatig gegenereerde beeldparen die zijn verkregen uit kunstmatige verplaatsingsvectorvelden (DVF's). Resultaten op CT-scans van de borstkas laten zien dat het opnemen van informatie met meerdere resoluties, en het hiërarchische gebruik via een LSTM hiervoor, leidt tot over het algemeen betere F1-scores, met minder misclassificaties in een goed afgestemde registratieconfiguratie. Het uiteindelijke systeem levert een nauwkeurigheid op van 87,1% en een gemiddelde F1-score van 66,4%, gemiddeld over twee onafhankelijke CT-scanonderzoeken van de borstkas.

## Discussie en vervolgonderzoek

Het werk dat in dit proefschrift wordt gepresenteerd, was gericht op het ontwikkelen van methoden voor het uitvoeren van beeldregistratie en het beoordelen van de kwaliteit van beeldregistraties.

In het voorgestelde RegNet in de hoofdstukken 2 en 3 hebben we een diepe convolutional neural network benadering gebruikt. Hoewel deep learning-methoden in segmentatietoepassingen veelbelovende resultaten hebben opgeleverd, bestaan er nog steeds verschillende uitdagingen in de registratietoepassingen. Het vinden van de optimale oplossing in conventionele segmentatiemethoden zoals level set en min cost (minimum van de kostenfunctie) is meestal iteratief, vergelijkbaar met conventionele

beeldregistratietechnieken. Bij conventionele beeldsegmentaties is het hoofdbeeld echter constant in alle iteraties en wordt de segmentatie in elke iteratie bijgewerkt. In tegendeel, bij de conventionele iteratieve beeldregistratie wordt bij elke iteratie een impliciete (of expliciete) herbemonstering van het vervormde bewegende beeld uitgevoerd. Blijkbaar is het voorspellen van de uiteindelijke transformatie in één keer nog steeds een uitdaging in op DL gebaseerde methoden. We merkten significante verbetering bij het gebruik van de meertrapsbenadering, waarbij ook een resampling werd uitgevoerd. Zoals vermeld in Tabel 2.2 is de gemiddelde TRE verbeterd van 3,80 mm naar 1,57 mm. Het is vermeldenswaard dat de registratiekwaliteit nog kan worden verbeterd door achtereenvolgens meerdere RegNets in de oorspronkelijke resolutie te gebruiken. Mogelijk kan een eenvoudig stopcriterium worden gebruikt, zoals het verschil in variatie tussen opeenvolgende transformaties of een meer complexe benadering, zoals reinforcement learning.

In Hoofdstuk 2 en 3 van dit proefschrift hebben we gebruik gemaakt van een gesuperviseerde transformatiebenadering. Heel wat artikelen stelden een niet-gesuperviseerde transformatiebenadering voor [23, 24, 8]. Een van de voordelen van de niet-gesuperviseerde transformatiemethode is dat de trainingsgegevens volledig realistisch kunnen zijn, inclusief volledig realistische ground truth transformaties. Gewoonlijk wordt de training uitgevoerd met een eenvoudige ongelijkheidsmetriek zoals mutual information. Een mogelijk nadeel van methoden zonder supervisie is dus dat de transformatie van de ground truth niet bekend is en dat het getrainde network niet noodzakelijk beter is dan een conventionele iteratieve registratie met dezelfde metriek voor ongelijkheid. Aan de andere kant, in de gesuperviseerde transformatiebenadering, is de ground truth transformatie accuraat (niet noodzakelijk uniek aangezien registratie meestal een slecht gesteld probleem is). Integendeel, de transformatie en meestal een van de beelden (het vervormde bewegende beeld) is misschien niet helemaal realistisch. Hoewel kunstmatige gegevensgeneratie een potentiële manier zou kunnen zijn om betere prestaties te krijgen dan menselijke experts, zou dit op dit moment te idealistisch kunnen zijn, waar de huidige op deep learning gebaseerde registraties veel verder zouden kunnen worden verbeterd. Al met al, om de gesuperviseerde benaderingen te verbeteren, kan de noodzaak van een grote medische dataset met een ground truth voor transformaties sterk worden gevoeld. Dit kan worden gedaan door onderscheidende oriëntatiepunten of regiosegmentaties te annoteren.

In Hoofdstuk 4 hebben we een regressiebenadering voorgesteld om de registratiefout te voorspellen en in Hoofdstuk 5 hebben we de taak vereenvoudigd tot een classificatiebenadering. Elk van deze benaderingen heeft zijn eigen voor- en nadelen. Benadrukt moet worden dat bij een regressiebenadering de aanvaardbare marge van de regressiefout correleert met de ground truth waarde. Een genormaliseerd verlies zou dus kunnen helpen om de training beter te laten convergeren. In de

classificatiebenadering wordt dit probleem geëlimineerd door het beschouwen van een klasse met grote waarden zoals  $[6, \infty)$  mm. Er moet echter op worden gewezen dat bij het definiëren van de labels als correct,  $[0, 3)$  mm, slecht  $[3, 6)$  mm, en verkeerd  $[6, \infty)$  mm, de labels niet meer ordinaal zijn maar meer lijken op geordende labels. Dit betekent dat een verkeerde classificatie tussen het correcte en een verkeerde label erger is dan een verkeerde classificatie tussen twee aangrenzende labels, zoals correct en slecht. Dit criterium komt van nature niet voor in classificatiebenaderingen, maar kan worden opgelegd met behulp van hiërarchische classificatie. Over het algemeen is de classificatie moeilijker voor waarden dicht bij de grens. Het is bijvoorbeeld niet triviaal om een waarde van 2,99 in de correcte of de slechte klasse in te delen. Een oplossing zou kunnen zijn om zachte ground truth labels te gebruiken, bijvoorbeeld voor de waarde 2,99 kan de ground truth worden ingesteld op  $[0, 60, 0, 40, 0]$  voor respectievelijk de klassen correct, slecht en verkeerd. In de hypersferische prototypebenadering [131] zal de one-hot gecodeerde ground truth worden toegewezen aan een outputruimte. Op deze manier kunnen we *a priori* informatie geven over de labels en die gebruiken in de organisatie van de outputruimte. Zo kunnen de verkeerde en de slechte labels dicht bij elkaar liggen, terwijl de verkeerde en de correcte labels meer afstand kunnen hebben. Een andere interessante toepassing in het hypersferische prototype is om tegelijkertijd de regressie en de classificatie uit te voeren, zelfs in dezelfde outputruimte.

In Hoofdstukken 2, 3 en 5 gebruiken we kunstmatige datageneratie om een convolutional neural network te trainen om registratie of registratiefouten te leren. We stelden benaderingen voor “single frequency”, “mixed frequency” en “respiratory motion” voor om kunstmatig verplaatsingsvectorvelden te genereren. Een van de beperkingen van de bovengenoemde generaties is dat ze naar verwachting gevoelig zijn voor anatomische veranderingen zoals tumorgroei. Deze beperking kan mogelijk worden aangepakt door een nieuw type vervorming toe te voegen aan de kunstmatige trainingsgegevensstrategie, die dergelijke anatomische veranderingen nabootst. Over het algemeen kan de kunstmatige generatie verder worden verbeterd door meer realistische en complexe simulaties toe te voegen. Als er bijvoorbeeld een ribsegmentatie beschikbaar is in CT-scans van de borstkas, is het mogelijk om niet-rigide vervorming buiten de rib en rigide vervormingen binnen de rib uit te voeren. Het network kan potentieel de relatie tussen organen en de stijfheid van de vervormingen leren. Andere realistische vervormingen zoals glijdende beweging van de longen kunnen ook aan de trainingsbeelden worden toegevoegd.

Hoewel alle experimenten in dit proefschrift worden uitgevoerd met CT-scans van de borstkas, zijn alle voorgestelde methoden generiek en kunnen ze mogelijk ook worden toegepast op andere modaliteiten en anatomische locaties. In een soortgelijk onderzoek naar de registratie van intra-subject MR hersenbeelden, werd RegNet

getraind op MR-beelden van de hersenen en liet het veelbelovende resultaten zien [21]. Het gebruik van kunstmatige datageneratie in afbeeldingen met meerdere modaliteiten moet in de toekomst echter worden onderzocht, omdat mogelijk een benadering voor het transformeren van de intensiteit [132] nodig kan zijn.

### **Algemene conclusies**

Concluderend stelt dit proefschrift leergebaseerde methoden voor ten behoeve van medische beeldregistratie en kwaliteitsbeoordeling van beeldregistratie. Alle voorgestelde methoden zijn volledig automatisch en vereisen geen menselijke interactie. De voorgestelde RegNet-architectuur werd getest op CT-scanparen van de borstkas en behaalde resultaten vergelijkbaar met een conventionele B-spline registratiemethode. Het hiërarchische classificatieraamwerk om onjuiste uitlijning van registratie te detecteren met behulp van long short term memory convolutional neural networks (ConvLSTM) heeft veelbelovende resultaten opgeleverd. Alle deep learning-methoden die in dit proefschrift worden beschreven, hebben een uitvoeringstijd in de orde van seconden, wat aanzienlijk beter is dan conventionele methoden.