AWARD NUMBER: W81XWH-19-1-0376

TITLE: MRI Volumetrics for Risk Stratification of Vision Loss in Optic Pathway Gliomas Secondary to NF1

PRINCIPAL INVESTIGATOR: Robert A. Avery, DO

CONTRACTING ORGANIZATION: Children's Hospital of Philadelphia

REPORT DATE: July 2021

TYPE OF REPORT: Annual report

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
July 2021	Annual	1JUL2020 - 30JUN2021

4. TITLE AND SUBTITLE

MRI Volumetrics for Risk Stratification of Vision Loss in Optic Pathway Gliomas Secondary to NF1

5a. CONTRACT NUMBER
W81XWH-19-1-0376

5b. GRANT NUMBER
NF180067

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S):

Robert A. Avery, DO

E-Mail: averyr@chop.edu

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Children's Hospital, Philadelphia
Research Institute
2716 South Street
Philadelphia, PA 19146-2305

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Our overall objective is to develop and validate a novel quantitative MRI analysis software application that will enable: 1) identifying children who are at the greatest risk for vision loss from their NF1-OPG; and 2) making treatment decisions based on accurate evaluation of disease progression and response to therapy. Over the past year, we have completed Tasks 1, 2 and 3 (Aim 1) and Task 4-5 (Aim 2) while making good progression on the remaining tasks and Aim 3 as listed in the SOW. Accurate MRI-base volumetric measurements have been achieved in the CHOP and CNHS data cohorts. We have performed analysis of longitudinal changes in tumor volume and shape and are actively working on identifying the optimal set of MRI features that predict clinical outcomes (Task 6, Aim 2). Subject enrollment in the International NF1-OPG Natural History study is on track with over 185 subjects enrolled and 103 having completed the 1-year outcome, confirming we will have a complete data set to finish Aim 3 as planned. We have not made any significant changes to our study plan.

15. SUBJECT TERMS
NF1, optic pathway glioma, vision

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	21	**19b. TELEPHONE NUMBER** *(include area code)*

TABLE OF CONTENTS

Introduction

Nearly 20% of children with Neurofibromatosis type 1 (NF1) will develop a tumor of the visual system. The tumor, called an optic pathway glioma (OPG), causes irreversible vision loss leading to permanent disability in about 50% of children who have NF1, typically between 1 to 6 years of age. For unknown reasons, the other 50% of children with NF1-OPGs will not lose vision. Unfortunately, doctors do not have a good way of identifying which children will lose vision and when the vision loss will occur. Given this uncertainty, some children will sustain lifelong disability from their vision loss, even despite receiving treatment for their tumor, likely because treatment is started only after the loss of vision occurs. Also, for these exact same reasons, doctors may unknowingly treat NF1-OPGs that would have never caused vision loss. To address these clinical challenges, we will develop a novel quantitative magnetic resonance imaging (MRI) application that will accurately identify which children with NF1-OPGs will lose vision, thereby providing an opportunity to provide early treatment and preserve their vision.

Keywords: Neurofibromatosis type 1 (NF1); Optic pathway glioma (OPG); Visual Acuity (VA), Magnetic Resonance Imaging (MRI);

Accomplishments

◦ **What were the major goals of the project?**

Specific Aim 1 *Enable accurate MRI volumetric measurements of NF1-OPGs and reaffirm the relationship between these measures and vision loss in two independent cohorts (N=100, retrospective).*	**Timeline**	**July 2021 Technical Report Completion**	**CHOP**	**CNHS**
Major Task 1: Data acquisition	Months			
Subtask 1: Acquire and curate dataset D1 (N=50) from CNHS	1-3	Completed		Dr. Linguraru
Subtask 2: Acquire and curate dataset D2 (N=50) from CHOP	1-6	Completed	Dr. Avery	
Milestone(s) Achieved: Acquired multi-institutional data (N=100)	6	Completed	Dr. Avery	Dr. Linguraru
Local IRB/IACUC Approval	3	Completed	Dr. Avery	IRB already approved
Major Task 2: Enable accurate MRI-base volumetric measurements of NF1-OPG from new independent cohorts				
Subtask 1: Ensure robustness to data protocols	1-9	Completed	Dr. Avery	Dr. Linguraru
Milestone(s) Achieved: Accurate volumetric analysis of NF1-OPG (error $\leq$ 0.65 mm)	9	Completed	Dr. Avery	Dr. Linguraru
Major Task 3: Reaffirm the relationship between OPG volumetric measures and vision loss in the new independent cohorts				
Subtask 1: Relation between MRI Volumetrics and Visual Acuity	6-12	Completed	Dr. Avery	Dr. Linguraru
Milestone(s) Achieved: Good correlation between NF1-OPG volumes and vision loss (r-squared $\geq$ 0.60)	12	Completed	Dr. Avery	Dr. Linguraru

Specific Aim 2: *Enable MRI-based comprehensive assessment of longitudinal OPG changes and determine which features are*	**Timeline**	**July 2021 Technical**	**CHOP**	**CNHS**

associated with vision loss in a large cohort of children (N= 100, retrospective).		**Report Completion**		
Major Task 4: Data acquisition				
Subtask 1: Access dataset D3 (N=100) from the NF1-OPG natural history study	3-12	Completed	Dr. Avery	
Subtask 2: Transfer dataset D3 (N=100) to CNHS	3-12	Completed		Dr. Linguraru
Milestone(s) Achieved: Full access to dataset D3	12	Completed	Dr. Avery	Dr. Linguraru
Major Task 5: Enable MRI-based comprehensive assessment of longitudinal OPG changes				
Subtask 1: Quantify the size, shape and texture of NF1-OPG with MRI	9-18	Completed	Dr. Avery	Dr. Linguraru
Subtask 2: Quantify longitudinal MRI changes of NF1-OPG	12-21	80%Completed		Dr. Linguraru
Milestone(s) Achieved: Tool for MRI comprehensive assessment of longitudinal OPG changes	21	80% completed	Dr. Avery	Dr. Linguraru
Major Task 6: Design MRI-based prediction model of vision loss (PPV ≥ 90%)				
Subtask 1: Identify and evaluate the optimal set of MRI features that predict clinical outcomes	21-27	40% Completed	Dr. Avery	Dr. Linguraru
Milestone(s) Achieved: MRI-based prediction model of vision loss based on retrospective data	27	In progress	Dr. Avery	Dr. Linguraru
Specific Aim 3: *Prospectively validate the predictive model of vision loss in a cohort of subjects enrolled in a large NF1-OPG longitudinal study from 25 international NF1 clinics (N=50, prospective).*				
Major Task 7: Prospective data acquisition				
Subtask 1: Enroll new patients in dataset D4 (N=50) from the NF1-OPG natural history study	12-22	Completed	Dr. Avery	
Subtask 2: Assess patients in dataset D4 (N=50) at 1 year from enrollment	22-32	20% Completed	Dr. Avery	
Subtask 3: Transfer dataset D4 (N=50, at least two time points per patient) to CNHS	12-32	70% Completed	Dr. Avery	
Milestone(s) Achieved: Full access to dataset D4	32	In progress	Dr. Avery	Dr. Linguraru
Major Task 8: Enable MRI-based comprehensive assessment of longitudinal OPG changes				
Subtask 1: Validate the predictive model of vision loss at baseline	14-24	90% Completed	Dr. Avery	Dr. Linguraru
Subtask 2: Validate the predictive model of vision loss at 1 year	24-36		Dr. Avery	Dr. Linguraru

Milestone(s) Achieved: Prospectively validated MRI-based prediction model of vision loss (PPV ≥ 80%)	36		Dr. Avery	Dr. Linguraru

◦ **What was accomplished under these goals?**

SA1/Major Task 2: Enable accurate MRI-base volumetric measurements of NF1-OPG from new independent cohorts. We have now completed this task in full. As expected, we were able to achieve consistent volumetric analysis that was accurate on data from CNH and CHOP. This required making adjustments to our algorithm given the differences between Siemens and General Electric MRI platforms, and between the site protocols. We developed an MRI-harmonization technique and an image super-resolution method. The first, achieved image harmonization using an unsupervised deep learning architecture based on variational autoencoder with cycle generative adversarial networks, which learns a common space using an unpaired image-to-image translation network. The segmentation of the anterior visual pathway is simultaneously generated by a supervised learning strategy, as shown in **Figure 1**. In addition, we developed a self-supervised, multi-sequence super resolution technique with a deep network that computes non-local spatial attention from high-resolution T1 image features and applies the attention to low-resolution T2 image features. This technique illustrated in **Figure 2** allows us to increase the image resolution of T2 images (clinically acquired a low resolution) and perform segmentation using a combination of T1 an T2 sequences, which improves the performance of our methodology.

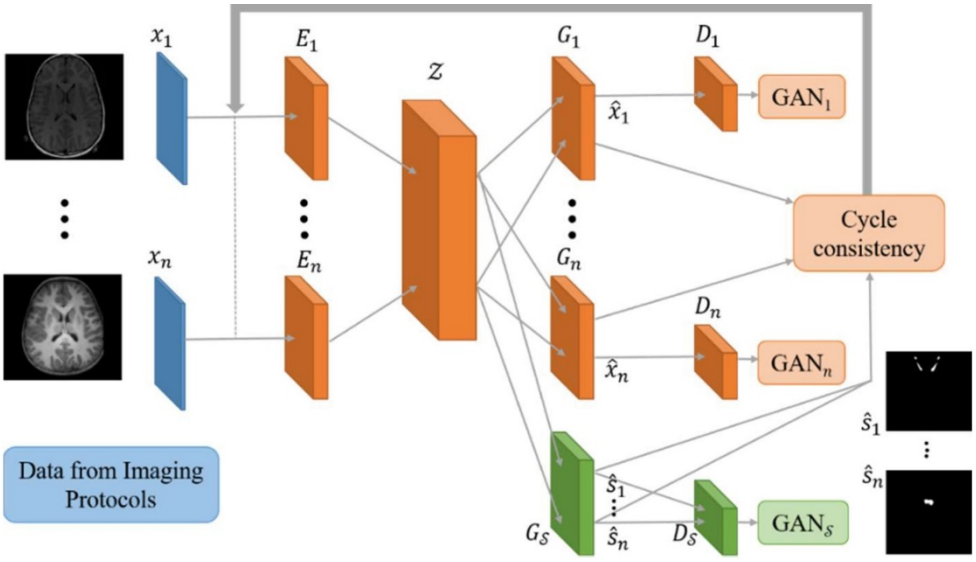

Figure 1. Schematic of our network for image harmonization.

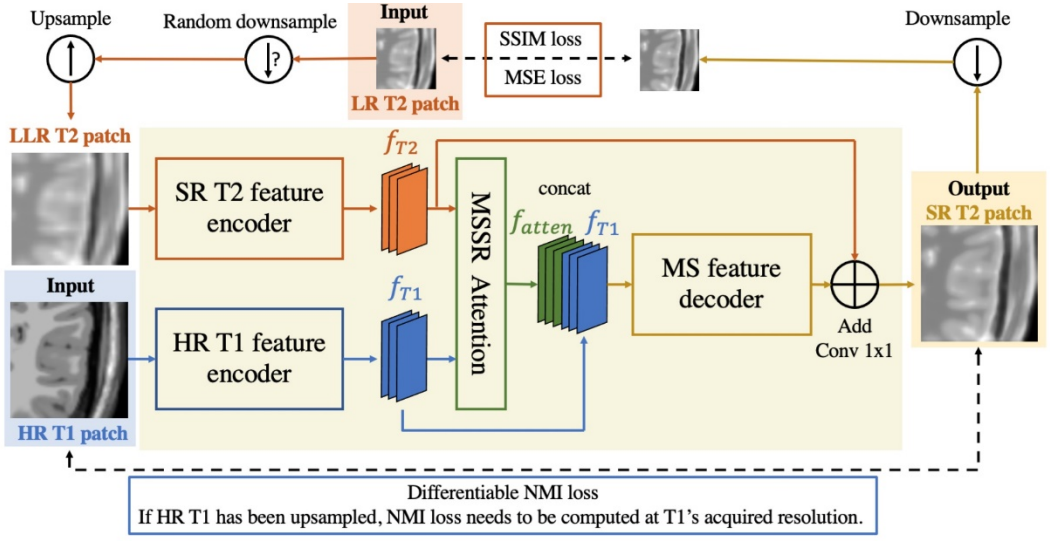

Figure 2. Super-resolution network architecture and training workflow. "LLR T2 patch" refers to randomly degraded low-resolution patch. At inference, the network takes upsampled low-resolution (LR) T2 and high-resolution (HR) T1 images instead of LLR T2

SA2/Major Task 5: *Enable MRI-based comprehensive assessment of longitudinal OPG changes and determine which features are associated with vision loss in a large cohort of children (N= 100, retrospective). Enable MRI-based comprehensive assessment of longitudinal OPG changes.* We have completed all manual segmentations as well as the volumetric size analysis for this large cohort derived from 2 centers with two different MRI platforms. In comparing the groups, the longitudinal analysis for the remaining portion of this aim will be achieved. We are currently determining which are the ideal boundaries for identifying focal (**Figure 3**) versus global changes over time (**Figure 4**). However, we have experienced some challenges with quantifying the shape and texture of NF1-OPG between these 2 different MRI platforms.

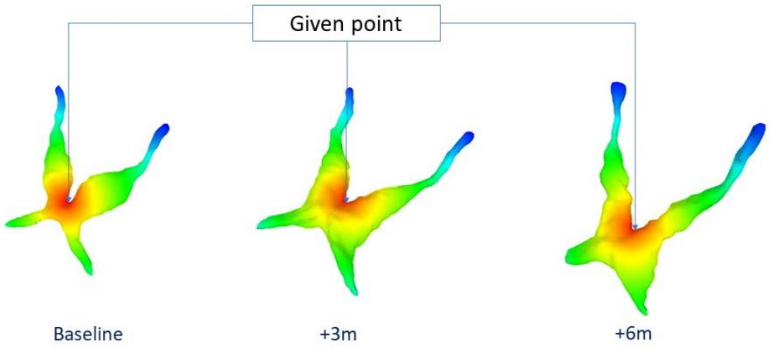

Figure 3 Geodesic distance from a given point

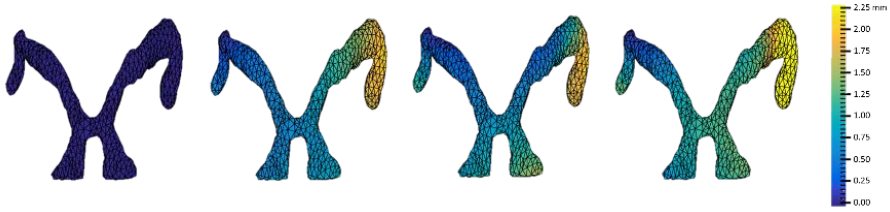

Figure 4. Temporal growth of AVP. Focal size/shape changes over time are quantified by the QI technology. The color map shows changes from the baseline.

SA2/Major Task 6 Design MRI-based prediction model of vision loss (PPV ≥ 90%) by identifying and evaluating the optimal set of MRI features that predict clinical outcomes. This task is on schedule and roughly 40% complete. Our preliminary analysis suggests that tumor volume remains the strongest predictor of vision loss, but we are exploring more dynamic and granular methods to determine which specific features of tumor shape and texture influence the model (**Figure 5**).

In is important to keep in mind that the analysis of tumor shape and texture is a very novel and cutting-edge approach to understanding clinical outcomes—which, to our knowledge has never been published in the field of NF1-OPG. Therefore, we are very thoughtful of our analysis approach and trying to avoid over or underfitting our models.

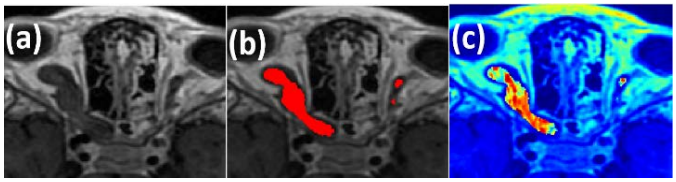

Figure 5. Texture analysis on NF1-OPG. (a) FLAIR MRI sequence; (b) automatic segmentation of NF1-OPG using QI technology; (c) quantitative texture features within OPG.

SA3/Major Task 7 Prospective Data acquisition: We have completed subtask 1 by enrolling the new patients in dataset D4 from the NF1 Natural History Study. Subject enrollment in the International NF1-OPG Natural History study is on track with over 195 subjects enrolled and 103 having completed the 1-year outcome, confirming we will have a complete data set to finish Aim 3 as planned. Seventy percent of subtask 2 has been completed with frequent data transfer as the D4 dataset continues to grow each month as sites submit data. Fifty percent of subtask 3 has been completed due to time needed for curating, quality assurance and preparation between CNH and CHOP.

SA3/Major Task 8 Enable MRI-based comprehensive assessment of longitudinal OPG changes: Subtask 1 is near completion as we are still refining our prediction model as noted above. Over the next few months this subtask should be completed and manuscript preparation will be begin. From there, we will quickly be able to initiate subtask 2 "Validate the predictive model of vision loss at 1 year." Once the final dataset is analyzed (Major Task 7, subtask 3), we will be able to determine if additional subjects with or without vision loss will

need to be included in the analysis to achieve statistical power. Given the robust enrollment numbers to date and how study enrollment has increased after the COVID19 pandemic research restrictions have been lifted, we are confident an appropriate sample size will be available to complete the proposed tasks and aims.

◦ **What opportunities for training and professional development has the project provided?**

We have provided technical and translational training to a postdoctoral fellow at Children's National Hospital who joined the project team in November 2019. The training covers machine learning techniques for quantitative imaging as well as translational applications of imaging for children with NF1-OPG. The fellow also received career development support to transition to an independent research position, including resources for publications, funding and mentorship. In addition, we have an intern from Princeton University, Brendan Wang, who is spending the summer of 2021 doing a virtual internship at Children's National and learning about open-source machine learning techniques for image analysis. Brendan works with Dr. Linguraru and his team and is also learning about pediatric brain tumors and the role of quantitative imaging in their management.

◦ **How were the results disseminated to communities of interest?**

Nothing to report.

◦ **What do you plan to do during the next reporting period to accomplish the goals?**

We have become much more facile with our online meetings and trouble-shooting over the past year, thus we will continue to follow this workflow as it has allowed us to accomplish our goals. If the COVID19 pandemic work standards change again (e.g., return to strict mitigation procedures) we are prepared to deal with these circumstances in order to achieve our study goals. On the other hand, if travel and in person meetings are determined to be safe, we will explore more traditional meetings between CNHS and CHOP.

IMPACT:

◦ **What was the impact on the development of the principal discipline(s) of the project?**

Nothing to report.

◦ **What was the impact on other disciplines?**

Nothing to report.

◦ **What was the impact on technology transfer?**

Nothing to report.

◦ **What was the impact on society beyond science and technology?**

Nothing to report.

CHANGES/PROBLEMS:

◦ **Changes in approach and reasons for change**

▪ There have been no significant changes to the objectives or scope.

◦ **Actual or anticipated problems or delays and actions or plans to resolve them**

▪ The COVID19 pandemic has slowed down different aspects of the study due to institutional safety guidelines. We created better work from home options with our laptops. We were able to utilize better data transfer modules from home that improve imaging transfer and data transfer timeliness and efficiency. Despite a decrease in typical in person interactions, we are operating at nearly our pre-pandemic efficiency.

◦ **Changes that had a significant impact on expenditures**

There has been a modest carryover in expenditures as Dr. Tor Diez was not hired until 11/18/2019.

◦ **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

There have been no significant deviations, unexpected outcomes or changes to the protocol. The institutional review board (IRB) approved this protocol on January 24, 2018 and there is no study expiration date as determined by the rules of the 2018 Common Rule (see prior communications with CDMRP).

◦ **Significant changes in use or care of human subjects**

Nothing to report.

◦ **Significant changes in use or care of vertebrate animals.**

N/A.

◦ **Significant changes in use of biohazards and/or select agents**

N/A.

PRODUCTS:

◦ **Publications, conference papers, and presentations**

▪ **Journal publications.**

Carlos Tor-Diez Antonio R. Porras, Roger J Packer, Robert A Avery, Marius George Linguraru. Multi-Platform MRI Quantification: Unsupervised MRI Homogenization: Application to Pediatric Anterior Visual Pathway Segmentation. In: Liu M., Yan P., Lian C., Cao X. (eds) Machine Learning in Medical Imaging. MLMI 2020. Lecture Notes in Computer Science, vol 12436. Springer, Cham. https://doi.org/10.1007/978-3-030-59861-7_19

▪ **Books or other non-periodical, one-time publications.**

Nothing to report.

▪ **Other publications, conference papers, and presentations.**

Avery, R.A. The NF1 OPG Consortium: An International Multi-Center Natural History Study of Newly Diagnosed Optic Pathway Gliomas Secondary to Neurofibromatosis Type 1" Children's Tumor Foundation Annual Meeting, New York, NY, June 2021

Clinical Predictors of Treatment-Refractory/Relapsed Optic Pathway Glioma in Patients with Neurofibromatosis Type 1. Kotch C. Avery RA, Bouffet E, de Blank P, Listernick R, Gutmann DH, Bornhorst M, Campen C, Li Y, Aplenc R, Liu GT, Fisher MJ. Children's Tumor Foundation Annual Meeting, New York, NY, June 2021

Avery, R.A. An International Multi-Center Natural History Study of Newly Diagnosed Optic Pathway Gliomas Secondary to Neurofibromatosis Type 1: Preliminary Clinical Characteristics" European Neurofibromatosis Meeting, Rotterdam, Netherlands, December 2020

Tor Diez, C. Porras, A.R. Packer, R. Avery, R.A. Linguraru, M.G.: Multi-Platform MRI Quantification: Deep Neural Network for Volumetric Analysis of Anterior Visual Pathway. In: Proceeding of Radiological Society of North America (RSNA) Annual Meeting (2020)

Linguraru, M.G.: Collaborative Efforts Are Essential for Deep Learning to Help Children’s Health and Rare Diseases . Invited Presentation at NVIDIA GPU Technology Conference (GTC 2021)

◦ **Website(s) or other Internet site(s)**

Nothing to report.

◦ **Technologies or techniques**

Nothing to report.

◦ **Inventions, patent applications, and/or licenses**

Mansoor, A. Linguraru, M.G. Avery, R.: Apparatus and Method for Segmenting Complex Structures. US/62/470,728 PCT/US2017/065562

◦ **Other Products**

Nothing to report.

PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

◦ **What individuals have worked on the project?**

Name: Robert Avery
Project Role: PI
Researcher Identifier: ORCID ID: 0000-0003-1453-7282
Nearest person month worked: 1.2
Contribution to Project: Dr Avery has coordinated the major activities of the project and has help gather and analyze the MRI and clinical data for specific aims 1, 2, and 3. Dr. Avery has worked closely with Dr. Linguraru (Qualified collaborator) through in person meetings and video conferences to ensure accuracy of all data acquisition and analysis. Dr. Avery supervised Ms Garcia (research coordinator).
Funding Support: N/A

Name: Marius Linguraru
Project Role: Qualified collaborator
Researcher Identifier: ORCID ID: 0000-0001-6175-8665
Nearest person month worked: 1.2
Contribution to Project: Dr. Linguraru has worked closely with Dr. Avery (PI) on coordinating the activities of the project. He leads the technical developments conducted at CNHS for quantitative imaging and volumetric analysis. Dr. Linguraru supervised Dr. Tor Diez and the technical analysis.
Funding Support: N/A

Name: Arielle Garcia
Project Role: Research coordinator
Researcher Identifier: N/A
Nearest person month worked: 1.8
Contribution to Project: Ms Garcia has been responsible for collecting and organizing subject's MRI, clinical data entry and regulatory compliance.
Funding Support: N/A

Name: Carlos Tor Diez
Project Role: Postdoctoral research fellow
Researcher Identifier: ORCID ID: 0000-0003-3339-5777
Nearest person month worked: 6.0
Contribution to Project: Dr. Tor Diez has provided technical expertise to the project in the areas of automated image segmentation and quantification and machine learning. He has been the primary developer of the software and methodology to assess optic pathway gliomas. Dr. Tor Diez has also supported the evaluation of the methodology and participates in dissemination processes.
Funding Support: N/A

◦ **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**

Nothing to Report.

◦ **What other organizations were involved as partners?**

Nothing to Report.

SPECIAL REPORTING REQUIREMENTS

As the project PI, Dr. Avery involved Dr. Linguraru as a "Qualified Collaborator" on this project. The above report will be duplicative between investigators. As instructed, the tasks are clearly outlined in the SOW table

for which investigator is responsible for the task. These assignments are applicable to all accomplishments listed in the above report.

APPENDICES:

We appended the following manuscript:

*Carlos Tor-Diez Antonio R. Porras, Roger J Packer, Robert A Avery, Marius George Linguraru. Multi-Platform MRI Quantification: Unsupervised MRI Homogenization: Application to Pediatric Anterior Visual Pathway Segmentation. In: Liu M., Yan P., Lian C., Cao X. (eds) Machine Learning in Medical Imaging. MLMI 2020. Lecture Notes in Computer Science, vol 12436. Springer, Cham. https://doi.org/10.1007/978-3-030-59861-7_19

Unsupervised MRI Homogenization: Application to Pediatric Anterior Visual Pathway Segmentation

Carlos Tor-Diez[1], Antonio R Porras[1], Roger J Packer[2, 3], Robert A Avery[4] and Marius George Linguraru[1, 5]

[1] Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Hospital, Washington, DC 20010, USA
[2] Center for Neuroscience & Behavioral Health, Children's National Hospital, Washington, DC 20010, USA
[3] Gilbert Neurofibromatosis Institute, Children's National Hospital, Washington, DC 20010, USA
[4] Division of Pediatric Ophthalmology, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[5] School of Medicine and Health Sciences, George Washington University, Washington, DC 20037, USA
ctordez@childrensnational.org

Abstract. Deep learning strategies have become ubiquitous optimization tools for medical image analysis. With the appropriate amount of data, these approaches outperform classic methodologies in a variety of image processing tasks. However, rare diseases and pediatric imaging often lack extensive data. Specially, MRI are uncommon because they require sedation in young children. Moreover, the lack of standardization in MRI protocols introduces a strong variability between different datasets. In this paper, we present a general deep learning architecture for MRI homogenization that also provides the segmentation map of an anatomical region of interest. Homogenization is achieved using an unsupervised architecture based on variational autoencoder with cycle generative adversarial networks, which learns a common space (i.e. a representation of the optimal imaging protocol) using an unpaired image-to-image translation network. The segmentation is simultaneously generated by a supervised learning strategy. We evaluated our method segmenting the challenging anterior visual pathway using three brain T1-weighted MRI datasets (variable protocols and vendors). Our method significantly outperformed a non-homogenized multi-protocol U-Net.

Keywords: MRI Homogenization, Brain MRI Segmentation, Deep Learning.

1 Introduction

Magnetic resonance imaging (MRI) provides anatomical information of soft tissues in a non-ionizing and non-invasive way. However, MRI intensities may lack interpretability because they do not have a tissue specific unit. MRI units are dependent on the provider and protocol criteria, thus challenging the application of image segmentation

and classification techniques. For example, the authors [1], obtained an error of 12,50% classifying MRI without homogenization. Therefore, normalizing MRI intensities is critical. As presented in [2], there are two types of intensity correction approaches: based on the properties of the MRI machine, and based on the imaged object. The first is focused on finding the static field inhomogeneity by testing different coils or calibrating with phantoms. The second consists of giving a standard unit to observed tissues. Despite its difficulty, the general preference in the literature is for the second approach, since it does not need additional acquisitions and steers to a more general solution.

Most MRI intensity normalization methods are based on histogram-matching techniques [3–5]. These methods usually use expensive preprocessing steps to make the histogram mapping structurally meaningful, such as non-rigid registration or patch-based approaches [6, 7]. Alternatively, Shinohara et al. [8] proposed a list of principles for defining biologically interpretable units and defined an efficient pipeline procedure with several simple steps. Deep learning has also been recently introduced to the MR homogenization. In [9], they applied several networks based on U-Net [10] to segment different structures. These networks were used to learn a subject-based histogram mapping, leading to a non-parametric normalization MRI method. In [11], authors replaced the popular MRI bias field correction, i.e., ITKN4 [12], by an artificial neural network. Finally, Jacobsen et al. [13] presented an evaluation of four intensity normalization methods based on their segmentation performance using a 3D fully convolutional neural network. These methods lack a general function that maps different input domains to a common space.

A few recent methods have presented neural networks to perform unsupervised image-to-image translation [14–16]. Liu et al. [14] introduced the concept of shared-latent space into the unpaired translation, which is a common space between datasets. In [15], they presented an efficient network called CycleGAN that avoids the common space by introducing cycle consistency. Using generative adversarial networks (GAN), this approach maps one image to another domain and re-maps this estimated image to the original space. In the UNIT architecture [16], the authors extended [14] by adding a variational autoencoder (VAE) that helped learning the shared-latent space. In the medical image context, these techniques were used to synthesize images from one modality or sequence to another, such as synthesizing CT from and to MRI [17, 18] or T1- and T2-weighted MR images [19]. However, these works did not optimize the imaging protocol to a particular task.

Nevertheless, the amount of harmonization methods for diffusion-weighted MRI (dMRI) was recently rising [20–23]. The application of machine learning techniques improve the traditional methods and cope the re-projection of spherical harmonics basis, a common representation for dMRI. In [23], authors implemented a network based on residual network, obtaining harmonized images. Moyer et al. [22] presented a VAE in order to split the inter-scanner variability from the subject anatomy and be able to reconstruct it to any scanner protocol, similarly to the first part of UNIT [16]. In [21],

they were focused on applying recent deep network blocks such as the multi-task learning to increase the algorithm effectiveness. Although all these methods present promising results, they lack for a specific application such as segmentation or registration. In addition, their application to conventional MRI scans, such as T1w or T2w, still have to be evaluated for future works. One example of this transition was presented in [29], reformulating the dMRI method [20] based on linear regression and empirical Bayes.

In this paper, we present an unsupervised network for general multi-protocol MRI homogenization to find the mapping of all images to a new and common protocol. We apply this novel concept to the challenging segmentation of the anterior visual pathway (AVP), which is thin (thickness around 2-3mm), long and amorphous as it contains the chiasm, optic nerves and optic tracks. In addition, the AVP has variable shape between subjects and patients with disease, and shows poor contrast with respect to adjacent tissues. Without surprise, there are only a few methods designed to segment the AVP and they have struggled to be accurate [24–26], in part because they lack a robust multi-protocol strategy for segmentation. This is particularly challenging when datasets are small, which is the case of AVP images and pediatric applications in general.

We use our method to quantify the AVP, which is critical for the diagnosis, prediction and treatment of children with optic pathway glioma who are at risk for vision loss. Given the rare but serious nature of these gliomas, imaging data are very limited as patients are seen by different institutions using variable MRI protocols. Hence, we propose a novel approach to synthetize a common domain (intuitively representing an optimal MRI protocol) from limited and variable imaging data. The limitations of data availability is handled by orienting the brain MRI from different protocols that can be reliably read by the human observer to a common direction that can be accurately interpreted by a deep learning network. Thus, our main methodological contribution is the creation of an unpaired image-to-image network to synthesize a new common protocol that is concurrently optimized to perform an image analysis task, which in our example is the segmentation of the AVP. Based on UNIT [16], we generalize the number of input datasets, leading to increased combinations of intermediate results and therefore of the complexity of the cycle consistency. We handle this complexity with a new loss function that leverages the extra number of inputs to perform a more efficient training. In addition, the segmentation sub-network is directly connected to the common space provided by the VAE, thus training for one single domain while the network has multi-domain inputs. The general nature of our approach makes it applicable to other image analysis tasks that benefit from consistency in imaging protocols.

2 Method

2.1 Unsupervised Image-to-image Translation (UNIT)

The UNIT deep-learning architecture [16] consists of two parts: one that creates a common space between image domains, and another that reconstructs each domain from

the common space. The first part allows building a shared-latent space that finds the common features of the input datasets. The reconstruction part is made of different GANs for each input image domain that generate the reconstruction of each input dataset space from the shared-latent space.

UNIT was designed as a method to translate two datasets, or two protocols in our case. Let $\mathcal{X}_1$and $\mathcal{X}_2$ be two MRI protocols. The goal of the unsupervised method is to find the joint distribution $P_{\mathcal{X}_1\mathcal{X}_2}(x_1, x_2)$ with marginals $P_{\mathcal{X}_1}(x_1)$ and $P_{\mathcal{X}_2}(x_2)$. This step assumes that a variable z in a shared-latent space $\mathcal{Z}$ can represent both inputs with the same code and that both can be recovered separately. Let E_1 and E_2 be the encoders that map $\mathcal{X}_1$ and $\mathcal{X}_2$ to $\mathcal{Z}$. Then, we can reconstruct $\mathcal{X}_1$ and $\mathcal{X}_2$ from $\mathcal{Z}$ by a combination of a decoder and a generator. Given the G_1 and G_2 networks that generate $\mathcal{X}_1$ and $\mathcal{X}_2$ from $\mathcal{Z}$, the architecture uses adversarial discriminators, D_1 and D_2, to optimize the loss function $\mathcal{L}$ by a minimax game:

$$\arg \min_{E_1,E_2,G_1,G_2} \max_{D_1,D_2} \mathcal{L}_{UNIT}(E_1, E_2, G_1, G_2, D_1, D_2). \quad (1)$$

At this stage, the architecture is equivalent to a classic autoencoder. Then, the architecture combines the sub-networks of different domains to create map functions that switch from one space to another. Let $F_{1\to2}$ and $F_{2\to1}$ be the mapping functions that map $\mathcal{X}_1$ to $\mathcal{X}_2$ and $\mathcal{X}_2$ to $\mathcal{X}_1$, respectively. These functions are defined by mixing the encoder and generator sub-networks of different domains, i.e. $F_{1\to2} = G_2(E_1(.))$ and $F_{2\to1} = G_1(E_2(.))$. By applying cycle consistency [15], the loss function terms $\mathcal{L}_{CC_1}$ and $\mathcal{L}_{CC_2}$ are computed such as $x_1 = F_{2\to1}\big(F_{1\to2}(x_2)\big)$ and $x_2 = F_{1\to2}\big(F_{2\to1}(x_1)\big)$, respectively.

Then, the loss function can be expressed as follows:

$$\begin{aligned}\mathcal{L}_{UNIT} = \; & \mathcal{L}_{VAE_1}(E_1, G_1) + \mathcal{L}_{GAN_1}(E_1, G_1, D_1) + \mathcal{L}_{CC_1}(E_1, G_1, E_2, G_2) \\ & + \mathcal{L}_{VAE_2}(E_2, G_2) + \mathcal{L}_{GAN_2}(E_2, G_2, D_2) + \mathcal{L}_{CC_2}(E_2, G_2, E_1, G_1).\end{aligned} \quad (2)$$

2.2 Generalized UNIT (G-UNIT)

The UNIT architecture is limited to two image domains, and to translating these domains into each other. However, MRI data often originate from multiple protocols, which would involve retraining the shared-latent space $\mathcal{Z}$ for each new data domain. We propose a more efficient alternative to expand the network by including n datasets and allowing the simultaneous training of $\mathcal{Z}$ for all protocols. Thus, in G-UNIT, we generalize the loss function from Eq. (2) as follows:

$$\mathcal{L}_{G\text{-}UNIT} = \sum_{i=1}^{n} \mathcal{L}_{VAE_i}(E_i, G_i) + \mathcal{L}_{GAN_i}(E_i, G_i, D_i) + \mathcal{L}_{CC_i}(E_1, G_1, \dots, E_n, G_n). \quad (3)$$

In addition, we introduce an initialization loss in the G-UNIT network. Since the image synthesis performance depend only on the loss cycle consistency term, the network does not optimize the latent space for the application as mentioned in [18]. If we consider that all our data are commonly oriented, the images from the same protocol should share the same contrast. To optimize the latent space for the application of the network, we add a term in the loss function that regularizes the image synthesis. Given an input image x_j acquired with the j-protocol, we implement the correlation coefficient loss function as follows:

$$\mathcal{L}_{init}(E_1, G_1, \dots, E_n, G_n) = -\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \frac{\left(\hat{x}_{ij} - m_{x_{ij}}\right)\left(x_j - m_{x_j}\right)}{\sqrt{\left(\hat{x}_{ij} - m_{\hat{x}_{ij}}\right)^2\left(x_j - m_{x_j}\right)^2}} \quad (4)$$

where $\hat{x}_{ij}$ is the estimated j-protocol for image x_i, $\hat{x}_{ij} = G_j(E_i(x_i))$, and $m_{\hat{x}_{ij}}$ and m_{x_i} are the mean intensity of $\hat{x}_{ij}$ and x_j, respectively.

2.3 Generalized UNIT for Segmentation (G-UNIT-S)

To incorporate the segmentation into G-UNIT and use the homogenized and optimized MRI protocol, we apply the segmentation directly to the shared-latent space $\mathcal{Z}$. This allows our method to train only one sub-network for segmentation, making it robust against inter-protocol variations and allowing the use of the entire data pool for training. For this purpose, we propose a GAN to map the shared-latent space $\mathcal{Z}$ to the segmentation space $\mathcal{S}$ and we implement a segmentation cycle consistency term to train the shared-latent space toward the optimal protocol representation. Since $\mathcal{S}$ and $\mathcal{Z}$ are trained simultaneously, we assume that the creation of $\mathcal{Z}$ is optimized for the generation of $\mathcal{S}$, i.e. $\mathcal{Z}$ contains the representation of the optimal protocol for segmentation. Let G_S be the generator of segmentations that maps from the common space $\mathcal{Z}$ and D_S the discriminator. The loss function of the GAN is:

$$\mathcal{L}_{\mathcal{S}}(G_S, D_S) = \mathcal{L}_{GAN_S}(G_S, D_S) + \mathcal{L}_{CC_S}(G_S, D_S). \quad (5)$$

Finally, the loss function of the G-UNIT-S method (Fig.1) is:

$$\mathcal{L}_{G\text{-}UNIT\text{-}S}(E_1, G_1, \dots, E_n, G_n, G_S, D_S) = \mathcal{L}_{G\text{-}UNIT}(E_1, G_1, \dots, E_n, G_n) + \lambda\, \mathcal{L}_{init}(E_1, G_1, \dots, E_n, G_n) + \mathcal{L}_{\mathcal{S}}(G_S, D_S) \quad (6)$$

where $\lambda = 1$ for the first epoch, and $\lambda = 0$ otherwise.

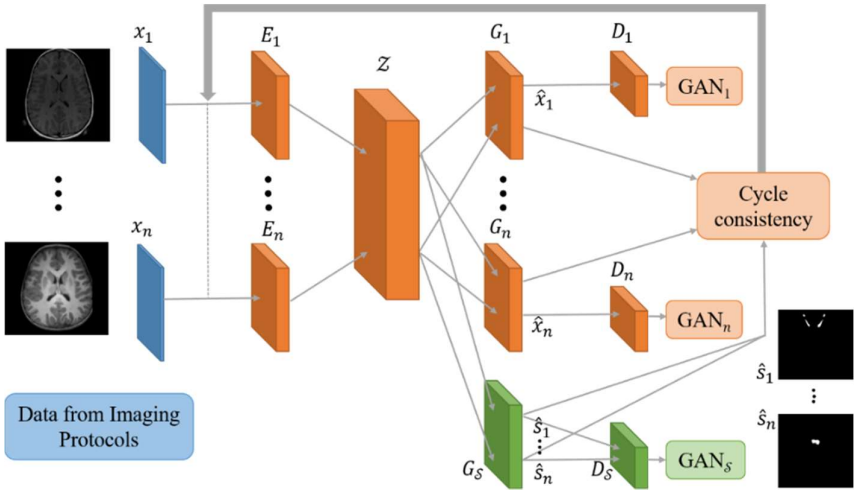

Fig. 1. Schematics of our proposed network G-UNIT-S. We consider n images $(x_1, \dots, x_n)$ from protocols $\mathcal{X}_1, \dots, \mathcal{X}_n$. $E_1, \dots, E_n$ are the encoders that map to the shared-latent space $\mathcal{Z}$. The generators $G_1, \dots, G_n$ reconstruct the image to the desired protocol. In addition, G_S generates the segmentation from $\mathcal{Z}$, i.e. for every protocol. With them, we can generate the reconstructions image $(\hat{x}_1, \dots, \hat{x}_n)$, other protocols $(\hat{x}_{1j}, \dots, \hat{x}_{nj})$ or segmentations $(\hat{s}_1, \dots, \hat{s}_n)$. $D_1, \dots, D_n$ and D_S are adversarial discriminators that regularize the results to be more realistic. All sub-networks are implemented as in [16].

3 Experiments and results

3.1 Data

Our approach is evaluated on multi-protocol brain MRI-data from three datasets. These datasets were acquired at three different centers: Children's National Hospital (CNH), Children's Hospital of Philadelphia (CHOP) and Children's Hospital of Colorado (CHC). Each center used a specific protocol and imaging device for all its acquisitions. More details about the imaging parameters are shown in Table 1.

Clinical Center	MRI Manufacturer	Acquisition Plane	TR/TE (ms)	Slice thickness (mm)	In-plane resolution (mm^2)
CNH	General Electric	Axial	600/10.5	0.60	0.41×0.41
CHOP	Siemens	Sagittal	1900/2.5	0.90	0.82×0.82
CHC	Philips	Coronal	8.3/3.8	1.00	0.94×0.94

Table 1. Properties of data used for our experiments grouped by acquisition center. TR- repetition time; TE- echo time.

3.2 Experiments and Results

As shown in Table 1, each dataset has a different acquisition plane and image resolution, which represents is an additional challenge to automate image analysis. We used rigid image registration to standardize the resolution between datasets, defining arbitrarily, the protocol with the highest resolution (CNH) as a reference. As a result, all images has a size of ($372 \times 324 \times 363$).

For training G-UNIT-S, datasets from each protocol have the same number of images, thus avoiding a skew in the generator bias. For our experiments, we used 18 3D T1-weighted MRIs, six from each clinical center, i.e., CNH, CHOP and CHC. The experiments are then performed by cross-validation using the leave-one-patient-out from each protocol strategy. To train our network, we used 100 slices for each 3D image. Therefore, we used 1,500 inputs and 25 epochs per each training in the cross-validation.

We used the Wasserstein distance with the gradient penalty variant [23] to train our network, which ensures stability to the training process. Data were postprocessed to keep only the largest component of the segmentation and remove small areas of false positives. We evaluated the reconstructed (after using cycle consistency) and the ground-truth images using the mean absolute square difference, and we used the Dice coefficient and the volume error to evaluate the segmentations. We compared the segmentation of the AVP from multi-protocol data using G-UNIT-S with the popular deep learning architecture for medical segmentation, U-Net [10]. U-Net does embed MRI homogenization, thus allowing us to assess the effect of MRI protocol homogenization for the image analysis task, i.e. AVP segmentation. Results are presented in Table 2. Note that the inter-observer variability was reported in an independent study with data from a single imaging protocol [26]; we expect it to be substantially lower on multi-protocol data. The significance of results was assessed at p-value ≈ 0.05 by using the Wilcoxon signed-rank test.

	U-Net	G-UNIT-S	Inter-observer variability
Dice coefficient	0.509 ± 0.223	0.602 ± 0.201	0.74 ± 0.08
Volume error (%)	$46.51 \pm 27.23\%$	$37.25 \pm 29.30\%$	N\A

Table 2. Results based on performing the Dice score and the volume error on estimated segmentation maps. These values correspond to the mean and standard deviation.

G-UNIT-S significantly outperformed the U-Net network using images from small datasets of variable acquisition protocols. Moreover, since the AVP does not have a stable contrast in MRI and its structure is thin and small, this improvement is still far from the inter-observer variability. A shortcoming of our approach is the limited data used for training the network. However, this is not uncommon in rare diseases where clinical consortia are distributed over multiple clinical centers. In addition, the segmentation of AVP is challenging due to the shape and appearance of this anatomical structure. Our experiments indicate that multi-protocol image homogenization is essential

for optimizing challenging image analysis tasks, and can significantly improve segmentation even for small datasets.

4 Conclusion

We presented a general deep network architecture that homogenizes different imaging protocols, and optimizes their common representation for an image analysis task. We demonstrated its application with using brain MRI and anterior visual pathway segmentation. Our network, called G-UNIT-S, was designed to homogenize any number of datasets, which is critical for the analysis of small datasets that are acquired by multiple clinical centers using different imaging devices and protocols. G-UNIT-S also incorporates an extension to segment the AVP based on the representation of a new, optimized image protocol. This extension can be expanded to general image analysis applications. Our results demonstrated that G-UNIT-S significantly improves an image analysis task over the popular U-Net. In future work, we will incorporate different modalities of MRI such as the T2-weighted and FLAIR to provide complementary information about the AVP anatomy.

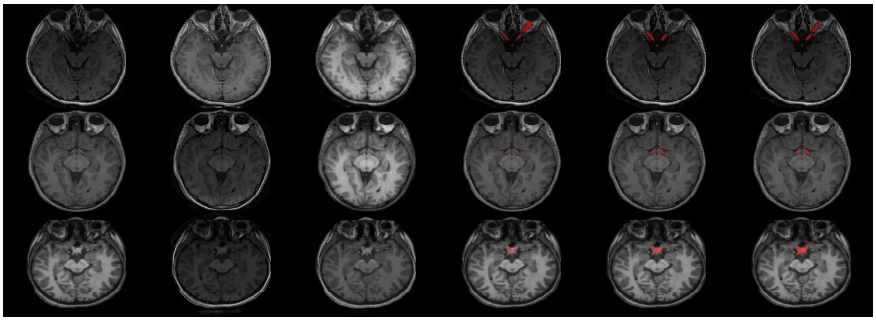

Fig. 2. Visual representation of results. From left to right: the original image, its transformation to other two protocols, the ground-truth segmentation and automated segmentations by U-Net and G-UNIT-S. From up to down: CNH, CHOP and CHC protocols.

References

1. Collewet, G., Strzelecki, M., Mariette, F.: Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magn Reson Imaging. 22, 81–91 (2004).
2. Vovk, U., Pernus, F., Likar, B.: A Review of Methods for Correction of Intensity Inhomogeneity in MRI. IEEE Transactions on Medical Imaging. 26, 405–421 (2007).
3. Nyul, L.G., Udupa, J.K., Xuan Zhang: New variants of a method of MRI scale standardization. IEEE Transactions on Medical Imaging. 19, 143–150 (2000).

4. Shah, M., Xiao, Y., Subbanna, N., Francis, S., Arnold, D.L., Collins, D.L., Arbel, T.: Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. Medical Image Analysis. 15, 267–282 (2011).
5. Sun, X., Shi, L., Luo, Y., Yang, W., Li, H., Liang, P., et al.: Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions. BioMedical Engineering OnLine. 14, 73 (2015).
6. Jager, F., Hornegger, J.: Nonrigid Registration of Joint Histograms for Intensity Standardization in Magnetic Resonance Imaging. IEEE Transactions on Medical Imaging. 28, 137–150 (2009).
7. Roy, S., Carass, A., Prince, J.L.: Patch based intensity normalization of brain MR images. In: 2013 IEEE 10th International Symposium on Biomedical Imaging. pp. 342–345 (2013)
8. Shinohara, R.T., Sweeney, E.M., Goldsmith, J., Shiee, N., Mateen, F.J., Calabresi, P.A., et al.: Statistical normalization techniques for magnetic resonance imaging. Neuroimage Clin. 6, 9–19 (2014).
9. Zhang, J., Saha, A., Soher, B.J., Mazurowski, M.A.: Automatic deep learning-based normalization of breast dynamic contrast-enhanced magnetic resonance images. arXiv:1807.02152 [cs]. (2018)
10. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 234–241. Springer International Publishing (2015)
11. Simkó, A., Löfstedt, T., Garpebring, A., Nyholm, T., Jonsson, J.: A Generalized Network for MRI Intensity Normalization. arXiv:1909.05484 [eess]. (2019)
12. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: Improved N3 Bias Correction. IEEE Trans Med Imaging. 29, 1310–1320 (2010).
13. Jacobsen, N., Deistung, A., Timmann, D., Goericke, S.L., Reichenbach, J.R., Güllmar, D.: Analysis of intensity normalization for optimal segmentation performance of a fully convolutional neural network. Z Med Phys. 29, 128–138 (2019).
14. Liu, M.-Y., Tuzel, O.: Coupled Generative Adversarial Networks. In: Advances in Neural Information Processing Systems 29. pp. 469–477. Curran Associates, Inc. (2016)
15. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks. Presented at the Proceedings of the IEEE International Conference on Computer Vision (2017)
16. Liu, M.-Y., Breuel, T., Kautz, J.: Unsupervised Image-to-Image Translation Networks. In: Advances in Neural Information Processing Systems 30. pp. 700–708. Curran Associates, Inc. (2017)
17. Wolterink, J.M., Dinkla, A.M., Savenije, M.H.F., Seevinck, P.R., van den Berg, C.A.T., Išgum, I.: Deep MR to CT Synthesis Using Unpaired Data. In: Simulation and Synthesis in Medical Imaging. pp. 14–23. Springer International Publishing, Cham (2017)

18. Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Xu, Z., Prince, J.: Unpaired Brain MR-to-CT Synthesis Using a Structure-Constrained CycleGAN. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. pp. 174–182. Springer International Publishing, Cham (2018)
19. Welander, P., Karlsson, S., Eklund, A.: Generative Adversarial Networks for Image-to-Image Translation on Multi-Contrast MR Images - A Comparison of CycleGAN and UNIT. arXiv:1806.07777 [cs]. (2018)
20. Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M.A., Ruparel, K., Roalf, D.R., Satterthwaite, T.D., Gur, R.C., Gur, R.E., Schultz, R.T., Verma, R., Shinohara, R.T.: Harmonization of multi-site diffusion tensor imaging data. NeuroImage. 161, 149–170 (2017).
21. Blumberg, S.B., Palombo, M., Khoo, C.S., Tax, C.M.W., Tanno, R., Alexander, D.C.: Multi-stage Prediction Networks for Data Harmonization. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. pp. 411–419. Springer International Publishing, Cham (2019)
22. Moyer, D., Steeg, G.V., Tax, C.M.W., Thompson, P.M.: Scanner invariant representations for diffusion MRI harmonization. Magnetic Resonance in Medicine.
23. Koppers, S., Bloy, L., Berman, J.I., Tax, C.M.W., Edgar, J.C., Merhof, D.: Spherical Harmonic Residual Network for Diffusion Signal Harmonization. In: Computational Diffusion MRI. pp. 173–182. Springer International Publishing, Cham (2019)
24. Noble, J.H., Dawant, B.M.: An atlas-navigated optimal medial axis and deformable model algorithm (NOMAD) for the segmentation of the optic nerves and chiasm in MR and CT images. Medical Image Analysis. 15, 877–884 (2011).
25. Yang, X., Cerrolaza, J., Duan, C., Zhao, Q., Murnick, J., Safdar, N., Avery, R., Linguraru, M.G.: Weighted Partitioned Active Shape Model for Optic Pathway Segmentation in MRI. In: Clinical Image-Based Procedures. Translational Research in Medical Imaging. pp. 109–117. Springer International Publishing, Cham (2014)
26. Mansoor, A., Cerrolaza, J.J., Idrees, R., Biggs, E., Alsharid, M.A., Avery, R.A., Linguraru, M.G.: Deep Learning Guided Partitioned Shape Model for Anterior Visual Pathway Segmentation. IEEE Transactions on Medical Imaging. 35, 1856–1865 (2016).
27. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved Training of Wasserstein GANs. In: Advances in Neural Information Processing Systems 30. pp. 5767–5777. Curran Associates, Inc. (2017)
28. Fortin, J.-P., Cullen, N., Sheline, Y.I., Taylor, W.D., Aselcioglu, I., Cook, P.A., et al.: Harmonization of cortical thickness measurements across scanners and sites. NeuroImage. 167, 104–120 (2018).

Acknowledgments

This work was funded by National Cancer Institute award UG3 CA236536 and Department of Defense award W81XWH1910376.