



HAL
open science

Leveraging noise and contrast simulation for the automatic quality control of routine clinical T1-weighted brain MRI

Sophie Loizillon, Stéphane Mabilie, Simona Bottani, Yannick Jacob, Aurélien Maire, Sebastian Ströer, Didier Dormont, Olivier Colliot, Ninon Burgos, Apprimage Study Group

► To cite this version:

Sophie Loizillon, Stéphane Mabilie, Simona Bottani, Yannick Jacob, Aurélien Maire, et al.. Leveraging noise and contrast simulation for the automatic quality control of routine clinical T1-weighted brain MRI. SPIE Medical Imaging 2024: Image Processing, Feb 2024, San Diego (CA), United States. hal-04674029

HAL Id: hal-04674029

<https://hal.science/hal-04674029>

Submitted on 20 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Leveraging noise and contrast simulation for the automatic quality control of routine clinical T1-weighted brain MRI

Sophie Loizillon^a, Stéphane Mabillet^a, Simona Bottani^a, Yannick Jacob^b, Aurélien Maire^b, Sebastian Ströer^c, Didier Dormont^{a,c}, Olivier Colliot^a, Ninon Burgos^a, and the APPRIMAGE Study Group*

^aSorbonne Université, Institut du Cerveau – Paris Brain Institute, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié-Salpêtrière, Paris, France

^bAP-HP, Innovation & Données – Département des Services Numériques, Paris, France

^cAP-HP, Hôpital Pitié-Salpêtrière, DMU DIAMENT, Dep. of Neuroradiology, Paris, France

ABSTRACT

The recent advent of clinical data warehouses (CDWs) has facilitated the sharing of very large volumes of medical data for research purposes. MRIs can be affected by various artefacts such as motion, noise or poor contrast that can severely degrade the overall quality of an image. In CDWs, a large amount of MRIs are unusable because corrupted by these diverse artefacts. Given the huge number of MRIs present in CDWs, manually detecting these artefacts becomes an impractical task. Therefore, it is necessary to develop an automated tool that can efficiently identify and exclude corrupted images. We previously proposed an approach for the detection of motion artefacts in 3D T1-weighted brain MRIs. In this paper, we propose to extend our work to two other types of artefacts: poor contrast and noise. We rely on a transfer learning approach, which leverages synthetic artefact generation, and comprises two steps: model pre-training on research data using synthetic artefacts, followed by a fine-tuning step, where we generalise the pre-trained models to clinical routine data relying on the manual labelling of 5000 images. The main objectives of our study were two-fold: to be able to exclude images with severe artefacts and to detect moderate artefacts. Our approach excelled in meeting the first objective, achieving a balanced accuracy of over 84% for the detection of severe noise and very poor contrast, which closely matched the performance of human annotators. Nevertheless, performance in the pursuit of the second objective was less satisfactory and inferior to that of the human annotators. Overall, our framework will be useful for taking full advantage of MRIs present in CDWs.

Keywords: Quality Control, Clinical Data Warehouse, Transfer Learning, MRI, Noise, Contrast, Motion

1. INTRODUCTION

Clinical data warehouses (CDWs) contain very large amounts of medical images collected from thousands to millions of patients, offering a unique opportunity to develop computational tools. Unlike research datasets that follow standardised acquisition protocols, images from CDWs are highly heterogeneous, originating from different hospitals, spanning multiple decades, and acquired using various machines without any homogenisation.

Magnetic resonance images (MRIs) are subject to many different types of artefacts that can compromise the overall quality of the image. In a previous study, we found that 25% of images tagged as 3D T1-weighted (T1w) brain MRIs in the CDW of Parisian hospitals were totally unusable, and that almost a third had a very low quality due to poor contrast, motion or noise artefacts.¹ As many images present in CDWs are corrupted by diverse artefacts, it is important to find an effective way to exclude them before conducting any study. Therefore, quality control (QC) appears as the crucial first step when dealing with images of CDWs.

At SPIE Medical Imaging 2023, we presented a transfer learning framework for the automatic detection of motion artefacts in 3D T1w brain MRIs from a CDW based on the simulation of motion artefacts on a research

*Members of the APPRIMAGE study group can be found at <https://www.aramislab.fr/apprimage>
Send correspondence to Sophie Loizillon: sophie.loizillon@gmail.com

dataset.² The results were promising, as we achieved highly accurate detection of severe motion artefacts, with a balanced accuracy exceeding 80%, which closely matches the performance of human annotators. While the use of synthetic motion to detect real artefacts in MRIs had already been studied,³⁻⁵ the simulation of noise or low contrast has mainly been used for data augmentation purposes⁶ rather than for artefact detection tasks.

In this paper, we propose to extend our previous work on motion artefacts to noise artefact and poor contrast by building a transfer learning framework for the automatic detection of these artefacts in 3D T1w brain MRI using a CDW. We generated synthetic poor contrast and noise artefacts in clean MRIs of a research dataset to train two new artefact-specific CNN classifiers. Our models were then generalised to clinical data with an effective transfer learning technique using 5000 manually labelled MRIs from a CDW.

2. MATERIALS AND METHODS

2.1 Dataset description

A publicly available research-oriented dataset was used to pre-train our CNNs with synthetic artefacts and one routine clinical dataset was exploited for transfer learning and validation.

The Alzheimer’s Disease Neuroimaging Initiative (ADNI) study is a multisite study of elderly individuals with normal cognition, mild cognitive impairment, or Alzheimer’s disease.⁷ The ADNI-1 phase included T1w MRIs acquired on 1.5 T scanners from different manufacturers (GE, Siemens, and Philips). We selected only artefact-free MRIs that passed the QC of ADNI and had no motion according to the IPMOTION score.

The clinical routine data comes from a large CDW containing all the T1w brain MRIs scanned in hospitals of the Greater Paris area (Assistance Publique-Hôpitaux de Paris). We used the same dataset as in our previous studies, where we randomly selected 5500 images that were acquired on various scanners (Siemens, GE, Philips and Toshiba).^{1,2} Artefacts in the MRIs were manually annotated by two annotators using a three-grade scale. Noise was scored as (0) when no noise was observed, (1) when there was noise present but it did not hinder structure identification, and (2) when severe noise prevented structure identification. Contrast was graded as follows: (0) for good contrast, (1) when grey and white matter were challenging to distinguish in certain regions, and (2) when the distinction between grey and white matter was difficult throughout the entire brain. Some of the 5500 images did not correspond to T1w brain MRI and were therefore not labelled with a score for the different artefacts (straight reject).¹ If the annotators labelled differently a given MRI, the consensus grade was chosen as the maximum of the two grades.

2.2 Proposed approach

We developed a transfer learning approach that enables the detection of noise artefacts and poor contrast on MRIs from the CDW. Our method leverages artefact simulation on the ADNI research dataset to achieve accurate detection of artefacts on routine clinical data.

2.2.1 Artefact Simulation

To simulate poor contrast in T1w MRIs, we aim to make it challenging to distinguish between white and grey matter in the brain. In this context, let I represent the artefact-free MRI image. We introduce a parameter β to control the non-linear gamma correction, which allows us to adjust the contrast. Applying the non-linear gamma correction to I , we obtain the corrupted MRI, denoted as $I_c(\beta) = I^{1/e^\beta}$. This approach was implemented using the `RandomGamma` transform function of TorchIO.⁸

To simulate noise in artefact-free MRIs, we aimed to replicate the inherent noise by modelling it as Gaussian distribution. For this purpose, we employed the TorchIO function `RandomNoise`, which simplifies the noise pattern.

2.2.2 Network architectures

To automatically detect artefacts, we implemented two 3D CNNs, denoted as Conv5FC3, composed of five convolutional blocks and three fully connected layers that proved successful in our previous work on motion detection.² Each convolutional block is made of a convolutional layer, a batch normalisation layer, a ReLU activation function and a max pooling layer. We used the ADAM optimiser, the weighted binary cross-entropy loss and a batch size of 6. Our implementation was done using Pytorch and the ClinicaDL software.⁹

2.3 Pretraining on synthetically corrupted research data

We conducted the pre-training on the ADNI research dataset of our Conv5FC3 models, where each of them is designed to detect a type of artefact: noise and bad contrast. For that purpose, we corrupted 192 artefact-free MRIs with synthetic artefacts of different natures and severity degrees. For introducing moderate and severe Gaussian noise into the MRIs, we utilised the `RandomNoise` transforms with standard deviation ranges of $\sigma=[5,15]$ and $\sigma=[15,25]$. To mimic the impact of moderate and severe bad contrast in our clean research MRIs with the `RandomGamma` function, we set the value of β ranges to $[-0.45, -0.3]$ and $[-0.2, -0.05]$, respectively. The ADNI corrupted dataset distribution for our two tasks is given in Table 1. It is important to note that to achieve robustness in artefact detection, our models were trained to handle cases where the label "No" for a specific artefact (e.g., "No noise") includes MRIs that are corrupted with other types of artefacts (e.g., poor contrast and/or motion). The images were split into training and validation using a 5-fold cross validation (CV). The separation between training and validation was made at the subject level to avoid data leakage. The same splits were used for all the tasks.

Table 1. Distribution of the label for the two artefact detection tasks: noise and bad contrast using the ADNI research dataset.

Task	Noise detection		Contrast detection	
Label	No noise	Noise (moderate and severe)	Good Contrast	Bad contrast (moderate and severe)
N° of MRIs	768	384	768	384

2.4 Application to clinical routine data using transfer learning

After pre-training our two artefact-specific models on the ADNI research dataset with synthetic artefacts, we proceeded with transfer learning techniques to adapt these models to clinical routine data. We fine-tuned our pre-trained models on 5000 images from the CDW on two specific target tasks: detecting severe (`Contrast01vs2`; `Noise0vs12`) and moderate (`Noise0vs1`; `Contrast0vs1`) artefacts. We tested our method on an independent test set of 500 clinical MRIs. However, it is important to consider the problem of limited annotations for severe noise (noise2: 16 MRIs). To address this issue, we adapted the severe noise detection task to focus on the `noise0vs12` task. This adaptation allowed us to leverage available images effectively and create a more balanced dataset. The CDW dataset distribution for our two tasks is given in Table 2. Before starting the experiments, we defined our training, validation and test sets by selecting the same MRIs as in.¹

Table 2. Distribution of the label for the two artefact detection tasks: noise and bad contrast using the clinical routine data of the CDW.

	Noise		Contrast	
Severe artefact detection	noise_0	noise_1/2	cont_0/1	cont_2
N° of MRIs	2382	1140	2317	1206
Moderate artefact detection	noise_0	noise_1	cont_0	cont_1
N° of MRIs	1464	533	937	652

3. RESULTS

The outcomes obtained using our proposed transfer learning framework on the independent clinical test set are presented in Table 3. For the detection of severe artefacts, the two new classifiers demonstrate high balanced accuracy (BA), which is comparable to that of manual annotators. The BA of the annotators is defined as the average of the BA between each rater and the consensus. We obtained a BA of 89.49% for the detection of poor contrast images (cont_2), just 1.5 percent points below that of manual annotators (BA: 91.03%). Similarly, we obtained a BA for severe noise detection (88.41%) that was almost identical to that of manual annotators (87.92%). For the detection of moderate contrast artefacts the classifier BA is low (72.70%) and lower than that of the raters (contrast: 84.87%). The moderate noise detection classifier achieved a very satisfying BA of 86.52%, less than 1 percent point lower than that of the manual annotators (noise: 87.22%).

We compared the performance obtained with our proposed transfer learning framework and when training the artefact-specific classifiers with the clinical data from scratch. Our transfer learning method achieved a gain of more than 1.2 and 4.5 percent points for respectively the detection of severe noise and poor contrast. The amount of improvement is less than for detecting severe motion in our previous work² but still high in the case of contrast. On the other hand, the improvement for noise was limited but one should note that the BA was already high when training from scratch for this task. A moderate but systematic improvement (2 percent points) was observed for the detection of moderate artefacts.

Table 3. Detection of noise and poor contrast artefacts on the CDW. For both the detection of severe and moderate artefacts, we report: the agreement between human raters and the consensus (manual annotations), results of the proposed approach (pre-training on synthetic artefact from research data and fine-tuning on CDW) and results when training from scratch on CDW. Our previous results for motion artefacts detection are also reported.² Cont: Contrast; Mov: Movement; BA: balanced accuracy.

		Severe artefact detection			Moderate artefact detection		
		BA	Specificity	Sensitivity	BA	Specificity	Sensitivity
Cont	Manual annotators	91.03	–	–	84.87	–	–
	Fine-tuning on CDW (proposed)	89.49	84.33	94.65	72.70	63.15	82.24
	Training from scratch on CDW	84.94	78.67	91.20	69.82	68.93	69.63
Noise	Manual annotators	87.92	–	–	87.22	–	–
	Fine-tuning on CDW (proposed)	88.41	82.88	94.81	86.52	77.62	95.41
	Training from scratch on CDW	87.23	79.45	95.00	84.79	81.60	87.98
Mov	Manual annotators	86.24	–	–	73.21	–	–
	Fine-tuning on CDW ²	84.52	83.67	85.37	62.61	52.00	73.23
	Training from scratch on CDW	73.75	49.58	97.93	58.93	28.81	89.05

4. DISCUSSION AND CONCLUSION

In this work, we extended our transfer learning framework for the automatic detection of motion to two new types of artefacts - noise and bad contrast - in 3D T1w brain MRIs from a CDW. After having pre-trained our CNNs to detect specific synthetic artefacts on images from a research database, we generalised our networks on clinical routine images by fine-tuning them on 5000 images of the CDW. We validated the proposed approach on 500 manually labelled clinical MRIs. We demonstrated the usefulness of generating artefacts in good-quality research images to improve their detection in routine clinical data. With our fine-tuning strategy, we effectively minimise the disparity that exists between research, where strict acquisition protocols are respected, and clinical data, which suffer from a lack of homogenisation of acquisition parameters. Our proposed transfer learning framework achieved very good results for the detection of severe artefacts with a BA of 89.49% and 88.41% for poor contrast and noise artefacts. These BA are nearly as good as that of the annotators (contrast: 91.03% and noise: 88.41%) and 1.2 and 4.5 percent points higher than when training the model from scratch. For detection of subtle artefacts, the performance was high for noise but moderate for poor contrast.

Overall, the present results demonstrate that our transfer learning approach is effective for detecting different types of severe artefacts (not only motion but also noise and poor contrast) in CDWs.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6).

The research was done using the Clinical Data Warehouse of the Greater Paris Hospitals. The authors are grateful to the members of the AP-HP DSN and URC teams, and in particular Stéphane Bréant, Florence

Tubach, Jacques Ropers, Pierre Rufat, Antoine Rozès, Camille Nevoret, Christel Daniel, Martin Hilka, Julien Dubiel, Cyrina Saussol and Rafael Gozlan. They would also like to thank the “Collégiale de Radiologie of AP-HP” as well as, more generally, all the radiology departments from AP-HP hospitals.

REFERENCES

- [1] Bottani, S., Burgos, N., Maire, A., Wild, A., Strer, S., Dormont, D., and Colliot, O., “Automatic quality control of brain T1-weighted magnetic resonance images for a clinical data warehouse,” *Medical Image Analysis* **75**, 102219 (2021).
- [2] Loizillon, S., Bottani, S., Maire, A., Ströer, S., Dormont, D., Colliot, O., and Burgos, N., “Transfer learning from synthetic to routine clinical data for motion artefact detection in brain T1-weighted MRI,” in [*Medical Imaging 2023: Image Processing*], **12464**, International Society for Optics and Photonics, SPIE (2023).
- [3] Mohebbian, M., Walia, E., Habibullah, M., Stapleton, S., and Wahid, K. A., “Classifying MRI motion severity using a stacked ensemble approach,” *Magnetic Resonance Imaging* **75**, 107–115 (2021).
- [4] Shaw, R., Sudre, C., Ourselin, S., and Cardoso, M. J., “MRI k-Space Motion Artefact Augmentation: Model Robustness and Task-Specific Uncertainty,” in [*Medical Imaging with Deep Learning - MIDL 2018*], (2018).
- [5] Duffy, B. A., Zhang, W., Tang, H., and Zhao, L., “Retrospective correction of motion artifact affected structural MRI images using deep learning of simulated motion,” in [*Medical Imaging with Deep Learning - MIDL 2018*], (2018).
- [6] Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., and Haworth, A., “A review of medical image data augmentation techniques for deep learning applications,” *Journal of Medical Imaging and Radiation Oncology* **65**(5), 545–563 (2021).
- [7] Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack, C. R., Jagust, W. J., Shaw, L. M., Toga, A. W., Trojanowski, J. Q., and Weiner, M. W., “Alzheimer’s Disease Neuroimaging Initiative (ADNI),” *Neurology* **74**(3), 201–209 (2010).
- [8] Pérez-García, F., Sparks, R., and Ourselin, S., “TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning,” *Computer Methods and Programs in Biomedicine* **208**, 106236 (2021).
- [9] Thibeau-Sutre, E., Díaz, M., Hassanaly, R., Routier, A., Dormont, D., Colliot, O., and Burgos, N., “ClinicaDL: An open-source deep learning software for reproducible neuroimaging processing,” *Computer Methods and Programs in Biomedicine* **220**, 106818 (2022).