École polytechnique fédérale de Lausanne

Semester project spring 2023

Master in Computational Science and Engineering

Quality control on Super-Resolution Reconstruction (SRR) for Fetal Brain MRI

Author:
Paul Devianne

Supervisor:
Dr. Thomas Sanchez
Dr. Meritxell Bach Cuadra
Dr. Jean-Philippe Thiran

Contents

1 Introduction **1**
1.1 General Context . . . 1
1.2 State of the art . . . 2
1.3 Problem at hand . . . 2
1.4 Contributions . . . 2

2 Methods **3**
2.1 Super-Resolution Reconstruction Problem . . . 3
2.2 Niftymic Outlier Rejection technique . . . 4
2.3 NeSVoR SRR method and Outlier Rejection . . . 5
2.4 Quality Control (QC) scoring . . . 6
2.4.1 Niftymic Uncertainty Measure . . . 6
2.4.2 Quality control with NeSVoR algorithm . . . 7

3 Experiments and Results **8**
3.1 Dataset . . . 8
3.1.1 Quality ratings . . . 8
3.2 Results . . . 9
3.2.1 Outlier Rejection Study . . . 9
3.2.2 Low-resolution quality control experiments . . . 11
3.2.3 Uncertainty Maps . . . 13
3.2.4 Super-resolution quality control experiments . . . 14

4 Discussions **16**
4.1 Significance of Metrics for Quality Control: Reliability, Complementarity, and Limitations . . . 16
4.2 Functionality of the Uncertainty Quantification maps . . . 17

5 Conclusion **17**

Abstract

SRR algorithms aims to reconstruct High-resolution volume of the fetal brain from Low-resolution MRI acquisitions. Evaluating the performance of these algorithms is crucial for assisting clinical diagnosis. The Quality Control on the algorithms is done by building adequate uncertainty metrics for the algorithms. The algorithms make use of outlier-rejection techniques, to correct potential artefacts from the MRI data due to fetal motion during the acquisition. The outlier-rejection methods are quantifying the amount of information discarded for the final reconstruction. Evaluating the parameters from these methods corresponds to building uncertainty metrics for the algorithm. Uncertainty maps of the reconstructed brains are built from the metrics. They aim at quantifying the uncertainty areas in the reconstructed fetal brain. Correlating the uncertainty findings with expert ratings evaluates the performance of the metrics on ground-truth data. Specific metrics displayed predictive power on the quality of the reconstructed fetal brain. However, lack of complexity in the quality ratings of the data as well as in the different metrics still retains high-performance uncertainty quantification.

Keywords: Fetal brain MRI, SRR algorithm, Quality Control, Uncertainty Quantification.

1 Introduction

1.1 General Context

MRI is being used more frequently to visualize the fetus and pregnancy structures in great detail. However, radiologists and clinicians only have access to Low-resolution MRI images along with more commonly used ultrasound (US) imaging. [8] The ultrasound imaging is used as a primary step for prenatal diagnosis of the fetal brain. If complementary information is needed, fetal brain MRI can be performed. T2-weighted MRI are used for enhancing structures mainly made of fluids like the fetal brain. [2]

Fetal MRI involves interactive scanning of the moving fetus using fast sequences, with Single-shot Fast Spin-Echo (SSFSE) T2-weighted imaging being commonly used. Advanced gestation brings significant changes in volume and T2-signal of fetal organs like the lungs and kidneys [2]. The fast-sequence technique allows us to freeze the in-plane-motion, making the final acquisition more robust to a possible subject motion.[9] The acquired MRI images are 2D-thick slices of the studied volume from a specific point of view. The slice thickness can range from 3 to 5 millimeters while in-plane resolution is inferior to 1 mm [**savio**]. This creates anisotropy in the acquired volume stack of slices. As for adult MRI, the orientation of the acquisition is manually selected to match the anatomical brain axis of the fetus. Yet, the supervision of a doctor during the acquisition is not standard for other procedures

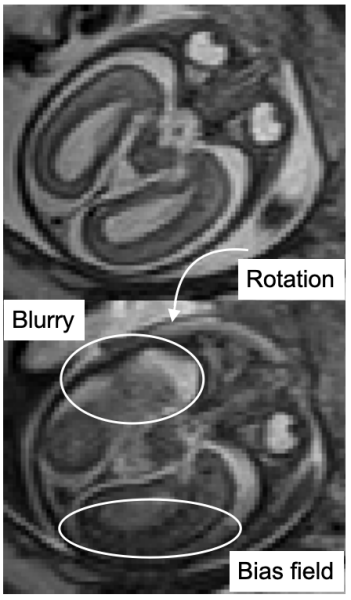


Figure 1: Neighbouring slices of CHUV subject MRI. The bottom slice is corrupted by artefacts.

Nevertheless, as the subjects can not be immobilized, the calibration might match the axis during the whole period of acquisition. This would generally induce two types of artefacts: inconsistencies between neighbouring slices, and corruption within the slice itself with bias field, signal drops or other possible alterations (see example on Figures 1 and 2) [5]. More generally, this motion can induce artefacts in the acquired MRI slices which, on top of some other factors, might alter the image quality . This corruption must be corrected if the slices are intended for clinical diagnosis.

In order to ease the fetal MRI analysis, a high-resolution (HR) isotropic 3D volume is needed. A Super-Resolution Reconstruction (SRR) algorithm is able to reconstruct such HR volume from low resolution input stacks of MRI slices [7, 5, 3, 1]. To start with, the localization and segmentation of the fetal brain are necessary to remove irrelevant information like maternal tissue that we need to removed for reconstructing the fetal brain Multiple artefacts are induced on the initial MRI images, like bias fields, signal drop, etc. The algorithms treat the artefacts by discarding or reducing the effect of the image elements associated to it. The MRI images are obtained using an interleaved acquisition of the slices. This technique is originally used to avoid slice cross-talk artefacts [5]. Yet, it allows neighbouring slices to present inconsistent position and orientation of the fetal brain. The resulting motion need to be corrected by the algorithm. The input LR slices are registered in a common spatial representation. This allows geometrical transformations on the slices to perform motion correction. The motion also induces different types of artefacts such as blurriness, signal drops, etc. The pre-processing and motion correction steps are not perfect. Then, even after the steps, slices still contain artefacts can hurt the reconstruction. Discarding these corrupted slices constitute the outlier rejection part of the algorithm. Although removing outliers ensures reliability of the input data, it can sometimes promotes the removal of additional information for the final reconstruction. The outlier rejection is under the following trade-off: removing slices that could hurt the reconstruction quality but removing too many slices would result in a loss of information. Quantifying this loss leads to a more complete understanding of the input slice treatment for the complete reconstruction. More generally, Quality Assessment and Quality Control (QA/QC) of the SRR algorithms aims at predicting how well the SRR performs. It should localize and quantify the confidence in the 3D reconstructed fetal brain.

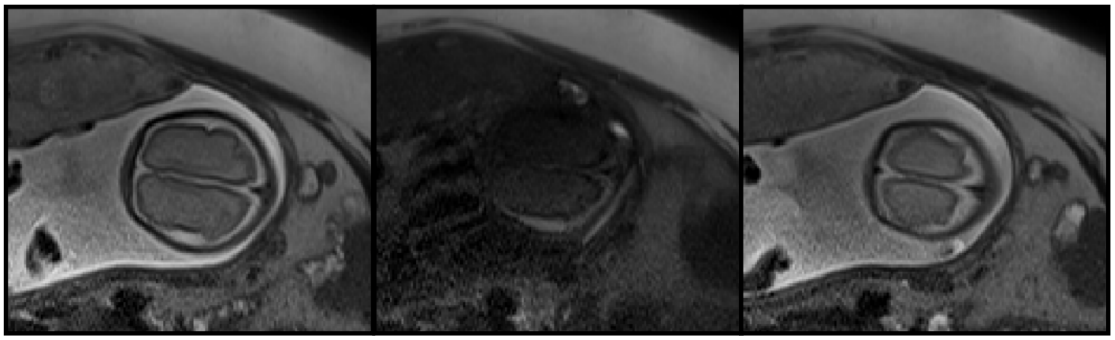


Figure 2: Neighbouring slices of CHUV subject MRI. The middle slice displays signal drop artefact.

1.2 State of the art

Within the fetal brain MRI research domain the QA/QC of the SRR algorithm has been studied through the correlation of the quality of the input low-resolution (LR) stacks with the SRR performance [4],[10]. The direct evaluation of the uncertainty, regarding the high-resolution (HR) volume output of SRR algorithms, has been indirectly studied through the performance evaluation of the algorithms [7, 6, 12, 11, 13, 3].

1.3 Problem at hand

The implementation of QC for SRR algorithms is essential for several reasons. Clinical use of reconstructed 3D volume requires a measure of the accuracy and reliability. Having information about the quality of the reconstruction enables medical practitioners to gauge the level of trust they can place in the results for making complementary diagnostics. This ensures that the reconstructed images are dependable and can aid in making informed medical decisions. This information is crucial for ensuring the accuracy and reliability of the reconstructed images in clinical settings. In order to assist clinical diagnosis, an uncertainty measure needs to be automated and coupled with the final reconstructed volume.

However, evaluating the uncertainty of the SRR algorithms can be challenging. To start with, one should expect to estimate the uncertainty from the different variables of the reconstruction. Although they are trained to correct these inconsistencies in the acquisition, the optimal reconstructed volume they converge to is not necessarily the most readable reconstructed volume for clinicians. The optimal volume can be different for the algorithm and the clinician. This can make the algorithm parameters deviate from a reliable uncertainty measure.

Secondly, the evaluation of the uncertainty from the algorithm perspective needs to match expert ratings of the same images.

1.4 Contributions

This paper presents contributions to the field of super-resolution reconstruction (SRR) of fetal brain MRI, specifically focusing on the evaluation of Niftymic [3] and NeSVoR [13] algorithms for high-resolution reconstructions of the fetal brain. These algorithms are recent SRR problem approaches. Niftymic uses the classical approach for solving the SRR problem [11], while NeSVoR is based on self-supervised learning. Most importantly, while Niftymic uses slice-wise outlier rejection, NeSVoR approach is a pixel-wise outlier-rejection. The primary objective is to assess the performance of these algorithms by quantifying the amount of information utilized at each location within the 3D reconstructed volume. ques to discard the corrupted data.

In our work, we endeavor to identify areas within the final volume that exhibit a reduced amount of information. These regions effectively form uncertainty quantification maps, which allow us to pinpoint locations where the reliability of the reconstructed volume may be uncertain. Building these uncertainty maps is a core aspect of this study.

Moreover, we aim to investigate the predictive capability of these uncertainty quantification maps regarding the quality of the reconstructed fetal brain volume. This evaluation is achieved by comparing the uncertainty maps with expert quality assessments of the reconstructed volume. Additionally,

we explore whether the uncertainty maps can be predictive of the quality of the low-resolution input slices, enabling us to gain insights into the reliability of the acquired data.

Predictive power of the uncertainty measure is evaluated from the correlation with expert assessment ratings of the LR input stacks and HR reconstructed volume. We aim at predicting expert ratings from specific metrics, building reliable predictors of the reconstruction uncertainty.

2 Methods

2.1 Super-Resolution Reconstruction Problem

Super-resolution reconstruction from 2D images to high-resolution volume is a challenging problem particularly when incorporating regularization techniques. In this subsection, we provide a general mathematical introduction to the problem of super-resolution reconstruction with TV regularization, outlining key concepts and mathematical formulations.

The super-resolution reconstruction problem can be mathematically described as follows: Given a set of observed low-resolution 2D images $\{y_i\}$ where $i = 1, 2, ..., N$, our objective is to estimate the corresponding high-resolution 3D volume x, which is not directly observable. The observed low-resolution 2D images y_i are related to the unknown high-resolution volume x through a degradation model that incorporates the effects of subsampling, motion and bias. Mathematically, this degradation model can be expressed as:

$$y_k = A_k \cdot x + e_k \tag{1}$$

where A_k represents the forward model, specific to each low-resolution image y_k, and e_k represents the additive noise associated with the k^{th} observation. The goal of super-resolution reconstruction is to recover the high-resolution image X given the set of observed low-resolution images $\{Y_i\}$ and the corresponding transformation $\{A_i\}$.

To solve the mathematical problem described in Eq. 1, we estimate the targeted volume through a maximum a-posteriori formulation:

$$x = \arg_{x \geq 0} \min \Big[\sum_k \frac{1}{2} \|y_k - A_k x\|^2 + \frac{\alpha}{2} \|\nabla x\|^2 \Big] \tag{2}$$

where k is the index of the MRI image used in a given set of sequential image acquisition. In this expression, the variable x represents the high-resolution volume to be estimated. The objective function consists of two terms. The first term $\sum_k \frac{1}{2} \|y_k - A_k x\|^2$, measures the data fidelity, where y_k represents the observed low-resolution image and A_k represents the corresponding imaging operator or degradation model for the kth image. The objective is to minimize the discrepancy between the estimated high-resolution volume x and the observed low-resolution images. The second term, $\frac{\alpha}{2} \|\nabla x\|^2$, represents the TV regularization term. Here, ∇x represents the gradient of the high-resolution volume x, and $||\nabla x||^2$ denotes its squared norm. The TV regularization promotes solutions that are piecewise smooth, encouraging sharp edges and reducing noise and artifacts in the reconstructed volume. The parameter α controls the strength of the regularization, determining the trade-off between data fidelity and regularization.

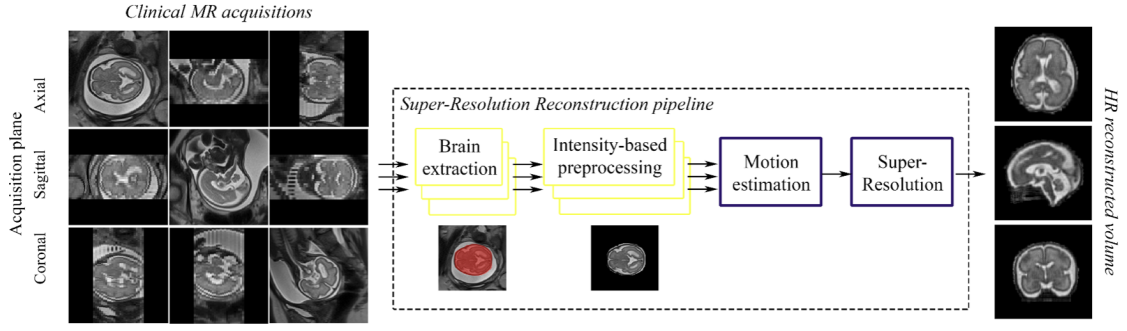


Figure 3: SRR algorithm pipeline presented by MIAL Laboratory from mialSRTK algorithm [11].

To solve the super-resolution reconstruction problem with TV regularization, optimization techniques are employed. The objective function is minimized with respect to the variable x by employing

suitable algorithms such as gradient-based methods, alternating minimization, or convex optimization methods. Recent advances present a self-supervised learning approach to the optimization problem. These algorithms iteratively update the high-resolution volume x to minimize the objective function and achieve an improved reconstruction.

Figure 3 summarizes the SRR pipeline. Firstly, the ROI is identified through localization and segmentation of the fetal brain in the input stacks. After pre-processing steps, the inter-slice motion is estimated to align neighbouring slices, matching the best way possible the delimited brain contours. Then, an interpolation is done from all the preprocessed stacks to the 3D HR volume. Although each SRR algorithm are based on different processing architecture, the different algorithms steps hold the same idea and goal as the ones presented above.

2.2 Niftymic Outlier Rejection technique

The Niftymic reconstruction algorithm [3] follows the super-resolution reconstruction (SRR) formulation depicted above, to map the stacks of 2D slices to the 3D volume. Additionally, Niftymic incorporates an outlier rejection mechanism, as an attempt of building an outlier-robust algorithm. This selection rule allows for discarding inconsistent data, ensuring that the final reconstruction is based on untainted input data.

To achieve robustness against outliers, the Niftymic algorithm performs multiple reconstruction cycles for the final reconstruction of the 3D targeted volume. The purpose of these cycles is to progressively discard the initial data that is too noisy to contribute productively to the quality of the final reconstruction.

In order to identify outlier slices, the Niftymic algorithm utilizes a similarity measure based on a first reconstructed volume obtained using all available slices. The idea is to map this reconstructed volume x back to the corresponding slice y_k using the forward model A_k, and evaluate their similarity. By comparing the reconstructed slices with the original data, outlier slices can be identified. For each reconstruction cycle, the volume is reconstructed, and then the outlier rejection step is performed. The rejection rule used in Niftymic is as follows: only the slices within a certain set, denoted as

$$\mathbf{K}_\beta = \big\{1 \leq k \leq K : \mathrm{Sim}(y_k, A_k x) \geq \beta\big\}, \tag{3}$$

are kept for the reconstruction in the next cycle. Here, β represents the rejection rate, and it is progressively increased through each cycle. This allows the algorithm to discard the most obvious outliers early on, improving the quality of subsequent reconstructed volumes. The similarity measure Sim is the usual Normalized Cross-correlation:

$$\mathrm{NCC}(I, J) = \frac{\sum_{x,y}(I(x,y) - \mu_I)(J(x,y) - \mu_J)}{\sqrt{\sum_{x,y}(I(x,y) - \mu_I)^2 \sum_{x,y}(J(x,y) - \mu_J)^2}} \tag{4}$$

By iteratively reconstructing the volume and refining the set of slices used in each cycle, the Niftymic algorithm effectively rejects outliers and enhances the robustness and accuracy of the final 3D reconstruction.

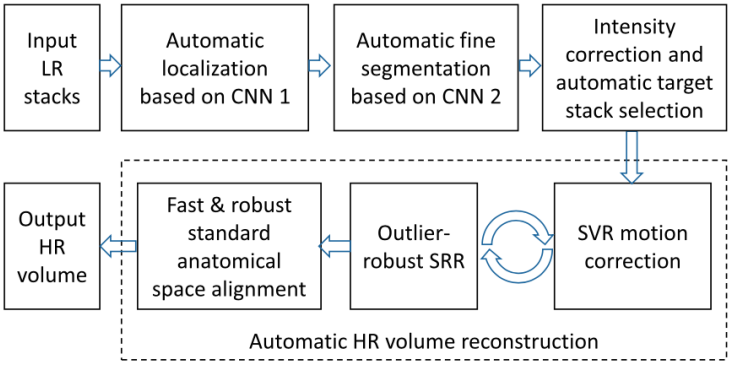


Figure 4: Niftymic algorithm pipeline [3].

Figure 4 summarizes the Niftymic algorithm pipeline. It combines the SRR formulation with an outlier rejection mechanism. Through multiple reconstruction cycles and an iterative rejection rule,

it selectively discards outlier slices. Robustness again outliers leads to an improved motion correction and a theoretical improvement of the reconstruction quality. We shall see that, although NeSVoR relies on the same idea of ill-posed problem for the SRR, the reconstruction and outlier-rejection present a different approach.

2.3 NeSVoR SRR method and Outlier Rejection

NeSVoR algorithm solves the SRR problem through the use of successive Multi-Layer perceptrons (MLP) [13].

The input slice intensities I_{ij} are directly mapped towards the 3D volume space, with i the index of the pixel of the j $^{\text{th}}$ slice. In the volume space Ω, the coordinates $x_{i,j,k}$ are transformed under a hash grid $\varphi(x)$, where k i the number of voxels for the reconstructed volume, dependant on the subject brain size as well as the desired resolution. The volume grid is then encoded as a neural network.

NeSVoR algorithm is a self-supervised model, grasping the underlying structure of the fetal brain by itself. It uses implicit neural models for the reconstruction as it is a powerful tool for generative models where we do not have prior knowledge of the high-resolution final volume. Two MLPs are used to predict the intensities of the final voxels and the variance of the reconstruction associated to it. The networks learn from the negative log-likelihood of Gaussian distribution:

$$\mathcal{L}_{i,j} = \frac{I_{ij} - \bar{I_{ij}}}{2\sigma_{ij}^2} + \frac{1}{2}\log\sigma_{ij}^2 \tag{5}$$

Here, the random variables that the networks are trying to learn are the mean pixel intensity $\bar{I_{ij}}$ along with the variance σ_{ij}^2 of the ground-truth data I_{ij}. NeSVoR model assumes that each intensity value from a slice is encoded in the following form:

$$I_{ij} = C_i \int_\Omega M_{ij}(x) B_i(x)[V(x) + \epsilon_i(x)] \tag{6}$$

$V(x)$ the target voxel intensities function, $B_i(x)$ the bias field, and $M_{ij}(x)$ the motion correction inverse transformation from volume to slice. Instead of integrating over the whole domain Ω, the spatial locations of the contributions of I_{ij} are described using a Gaussian point-spread function:

$$g(u;\Sigma) = \frac{1}{\sqrt{(2\pi)^3 \det(\Sigma)}} \exp\left(-\frac{1}{2}u^T\Sigma^{-1}u\right) \tag{7}$$

which gives the following form for the motion correction of the pixel i of slice j:

$$M_{ij}(x) = g(T^{-1} \circ x - p_{ij}; \Sigma). \tag{8}$$

Here p_{ij} are the spatial coordinates of the corresponding pixel. The idea behind these tricks is that the voxel function and associated transformation aims at producing the ground truth data I_{ij} through gaussian peaks at each point in the target domain of the volume. Using Monte-Carlo integration, the mean and variance of the intensity peaks are predicted as:

$$E[I] = \frac{C_i}{K}\sum_{k=1}^{K} B(x_{ijk})V(x_{ijk}) \tag{9}$$

$$\text{var}(I_{ij}) = \frac{C_i^2}{K}\sum_{k=1}^{K} M_{ij}(x_{ijk})B^2(x_{ijk})\sigma_i^2(x_{ijk}) \tag{10}$$

As multiple random variables are estimated through the different networks (see Figure 5), the uncertainty of the reconstruction can be evaluated from all of these different outputs. First, as one can observe from Eq. 5, the variance σ_{ij}^2 acts as an outlier-robust parameter. This parameters represents how well a pixel can help for the reconstruction. The model will give higher σ_{ij}^2 values for outliers. Higher values of this parameters will result in less impact in the reconstruction since the networks do not learn from the corresponding pixel. It plays the role of outlier removal pixel-wise. From the estimation of the mean and variance of resulting intensities with respect to the point spread function (PSF), another uncertainty measure can be made from these parameters. Indeed, the PSF

also induces variability in the reconstruction which should be taken into account when studying the uncertainty of the reconstruction. The motion correction and bias field operators have an impact on the final mean and variance of the voxels.

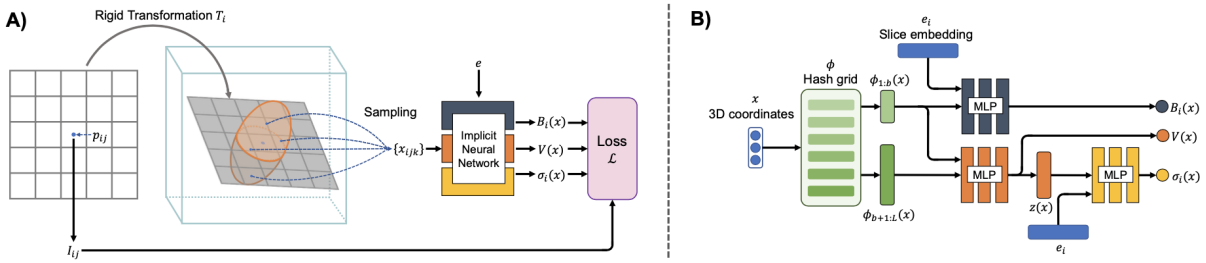


Figure 5: NeSVoR algorithm pipeline [13]. In A), the general pipeline from input pixel intensities to the loss function evaluation. In B), the subdivision of the different tasks between the MLPs.

Ultimately, the total variance associated with pixel I_{ij} is the sum of pixel-wise variance $\text{var}(I_{ij})$ and slice-wise variance defined by ν_i for which the effect on the reconstruction can also be lowered if considered necessary:

$$\sigma_{ij}^2 = \text{var}(I_{ij}) + \nu_i^2 \tag{11}$$

The details on the slice-wise outlier-robustness is not detailed in J.Xu and et al. paper. Then, NeSVoR outlier rejection is pixel-wise and slice-wise based.

Different outlier rejection techniques The removal of outlier for discarding useless information is built differently for the two SRR algorithms studied here. Niftymic algorithm evaluates similarity between *prior*- and *post*-reconstruction slices. Once detected, outliers are completely removed from the set of reconstructing slices. Although NeSVoR evaluate slice-wise variance for the computation of the pixel weight σ_{ij}^2, the more interesting approach is the pixel-outlier robust technique. The algorithms values/devalues pixels weight on the reconstruction based on their correlation with the initial intensities. This technique is a smooth outlier removal since the loss is a continuous functions of the parameter σ_{ij}^2. We shall see that this helps reducing the loss of useful information that could occur during outlier rejection steps.

2.4 Quality Control (QC) scoring

For the two studied algorithms the quality control is not performed similarly since the uncertainty measures are not similar. We present each uncertainty measure and the evaluations of the metrics attempted.

This work had access to expert ratings of LR input stacks and HR reconstructed volume. This allows the computation of correlations between uncertainty measurements to substantial ratings. The correlations serve as performance evaluation of the uncertainty metrics.

2.4.1 Niftymic Uncertainty Measure

The outlier rejection of Niftymic is an indicator of the amount of information used in the final HR reconstructed volume. More precisely, the number of slices participating in the reconstruction at specific location displays the amount of information available for the reconstruction of each voxel in the final volume. Computing the ratio of rejected slices quantifies the amount of information the algorithm had to discard for its reconstruction. The rejected slices are saved after the last reconstruction cycle, since they did not participate in the final HR volume. A first study consists of trying to correlate the quality of the LR stacks ratings with this ratio. We can use an absolute metric for each initial stack Y_i:

$$\text{AR}(Y_i) = |Y_i| - N_{rejected}(Y_i) \tag{12}$$

where $|\cdot|$ returns the number of elements, i.e. the number of slices in the stack Y_i, i the index of the stack, and $N_{rejected}(Y_i)$ the number of rejected slices for that stack. This measures the quantity

of slices that participated in the reconstruction for a specific stack. A relative ratio can also be an alternative:

$$\mathrm{RR}(Y_i) = \frac{N_{rejected}(Y_i)}{|Y_i|} \tag{13}$$

This gives out the ratio of rejected slices per stack which gives out information on the quality of the acquisition. A stack might contain many slices but utilizes only 30% of them. This would be detected by $\mathrm{RR}(Y_i)$ metric but not $\mathrm{AR}(Y_i)$. Comparing the metrics with LR rating metrics, we should expect that for a higher number of rejected slices, the ratings should be low. The AR metric is expected to correlate positively with the ratings. The RR is expected to correlate negatively with the ratings.

Then, for creating uncertainty maps, metrics should be a function of voxels in the final volume space. Two metrics can be evaluated for evaluating an uncertainty on the reconstructed volume. First, the absolute confidence denoted $\mathrm{AC}(x)$ is defined as:

$$\mathrm{AC}(x) = |Y(x)| \tag{14}$$

where $Y(x)$ is the set of slices that struck voxel x in the final reconstruction. This metric is an absolute measure. Indeed, comparing $\mathrm{AC}(x)$ between two subjects is consistent since a voxel struck by more slices for its final intensity value should have a higher accuracy. This measure is linked to the outlier rejection since more rejected slices would mean a lower $\mathrm{AC}(x)$ value. Yet, the $\mathrm{AC}(x)$ can present very low variation from the whole HR volume perspective. For this purpose, a second relative metric can also be defined $\mathrm{RU}(x)$. This metric is relative to each subject, it can be used for observing the uncertainty variation in the reconstructed volume but is not consistent for comparing subjects.

$$\mathrm{RU}(x) = |S| - \min(|Y(x)|, |S|) + 1 \tag{15}$$

where S is the set of stacks for the reconstruction. Here, the assumption is that the maximum number of slices involved in reconstructing a given voxel x. This ensures no overlapping from neighbouring slices. The higher the number of slices $|Y(x)|$, the lower the uncertainty. This metric is relative to the subject since different subject reconstructions use different number of stacks for the HR final volume. Comparing with HR volume quality ratings, we expect a positive correlation of AC with the ratings while it should be negative for RU.

To test the predictive power of these uncertainty measures maps, we use human-expert visually-made ratings of Low-resolution input stack of slices and High-resolution reconstructed volume. Once these uncertainty metrics built, one needs to correlate the results with the ratings. For the LR uncertainties performance evaluation, we simply compute the Pearson P, Spearman S correlation and r^2 value of the stack rating and the stack metrics as defined in Eq.12, 13. The results can be clipped to stay in between the q^{th} quantiles, removing potential outlier ratings/uncertainty measures. This technique will be indicated when applied.

For the evaluation of the HR volume uncertainty metrics, the metrics are averaged over the final volume of the subject since ratings of HR volume are only available for the complete volume not as volume maps like the uncertainties. Once averaged, we use the same correlation measures P, S, r^2 and potentially q.

2.4.2 Quality control with NeSVoR algorithm

Similarly as for Niftymic, different metrics are elaborated depending on if we wish to perform QC with the LR input stacks or the HR reconstructed volume.

As previously stated 2.3, the σ_{ij}^2 are an adequate representation of the amount of information the algorithms discard. The assumption made here is that, as for Niftymic, the more information discarded is linked to a higher uncertainty. Yet, there are multiple algorithm variables that could be correlated to the uncertainty. In total, 22 metric variables were tested for the experiments. A metric should be representative of complete stack uncertainty if we want to correlate its evaluation with the corresponding rating. The straightforward metric variables are listed here:

$$\mathbb{E}(\sigma_{ij}^2) \tag{16}$$

$$\mathbb{E}(\mathrm{var}(I_{ij}^2)) \tag{17}$$

$$\mathrm{var}(\sigma_{ij}^2) \tag{18}$$

$$\mathrm{Median}(\mathrm{var}(I_{ij}^2)) \tag{19}$$

The $\mathbb{E}(\cdot)$ and $\text{var}(\cdot)$ operators are evaluated over the whole stack. Eq. 16-19 metrics are expected to correlate negatively with the LR stack ratings as they are linked to the amount of outlier-pixels. Similarly to Niftymic case, one can evaluate a normalized metric, by dividing all σ_{ij}^2 values by the maximum value $\max_{i,j}(\sigma_{ij}^2)$ to highlight contrast for a specific subject.

For the HR reconstructed metric study, different metrics are defined. Indeed, σ_{ij}^2 variable is initially computed in the LR input space. The values are fixed for one specific pixel from a slice. The first step for building the sigma HR map is to interpolate and sum all values of σ_{ij}^2 onto their location in 3D final volume space Ω. To perform this interpolation, we used similar functions as for Niftymic algorithm, to aggregate all the sigma values to the output volume space. Once this operation has been performed, a high-resolution map of sigma values is obtained. It can then be compared with the HR fetal brain reconstructed volume. With this 3D $\sigma^2(x)$ map, one can define two uncertainty metrics for each voxel in the HR reconstructed volume. The absolute uncertainty metric is:

$$\mathbb{E}(\sigma^2(x)) \tag{20}$$

The $\mathbb{E}(\cdot)$ operator is evaluated over Ω, and $x \in \Omega$. The relative uncertainty metric is:

$$\frac{\mathbb{E}(\sigma^2(x))}{\max_x(\sigma^2(x))} \tag{21}$$

One would expect these metrics to correlate negatively with the ratings since a high value $\sigma^2(x)$ is still related to less information at voxel x.

3 Experiments and Results

3.1 Dataset

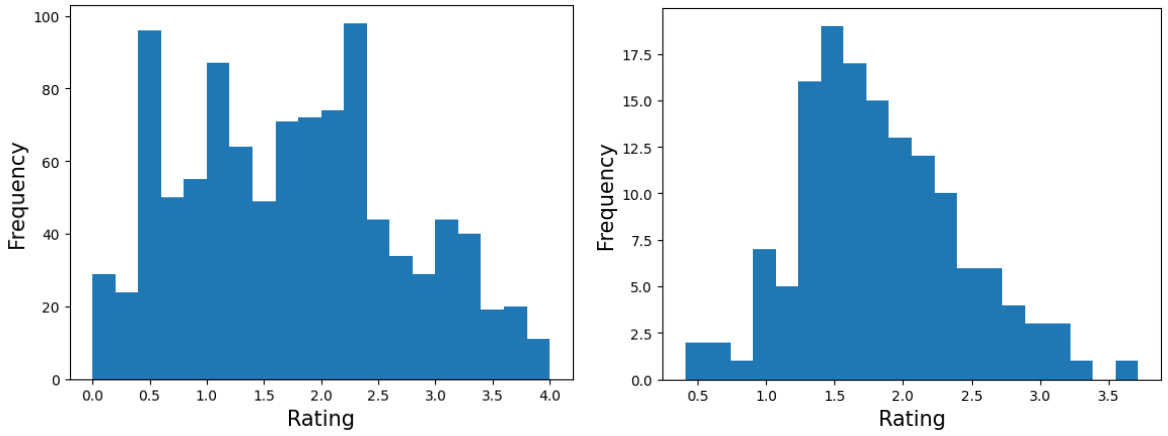
The SRR algorithm were performed on a 20 subjects dataset. Each subject is represented by a set of 4 to 14 stacks. The stacks were acquired at CHUV and made accessible for MIAL Laboratory. In-plane resolution and through-plane thickness of slices are respectively of 1 mm and 5 mm. The resolution parameter for the SRR algorithms is set to 0.8 mm.

Manually selected stacks are known to decrease the performance of the SRR algorithms. A manually outlier stack rejection is done by making a second reconstruction per subject. The *run-0* reconstruction uses all available stacks from a subject while *run-1* uses the manually selected stacks. Reconstructions can be unsuccessful. Then, some subjects only have one of the two runs. For Niftymic SRR algorithm, 29 reconstructions were performed successfully. For NeSVoR SRR algorithm, 31 reconstructions were performed successfully.

3.1.1 Quality ratings

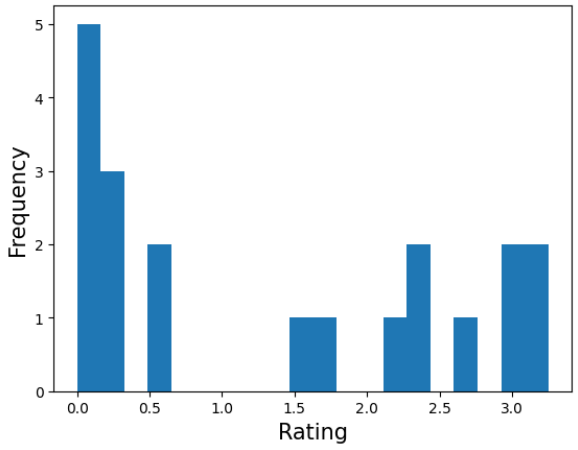
Dr. Meritxell Bach Cuadra conducted manual ratings on the Low-Resolution input stacks and High-Resolution reconstructed volume used in this work. The ratings ranges from 0 to 4, 0 being minimum quality and 4 the maximum. These ratings are considered as ground-truth data. Analysing them gives new insights for discussions.

The ratings over the whole dataset are summarized by the histograms on Figure 6.

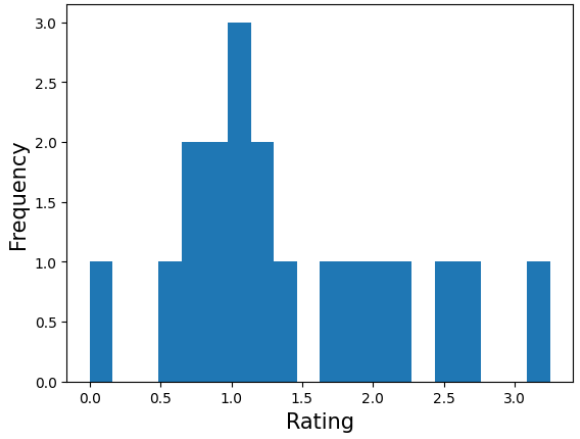


(a) Ratings LR over all stacks and subjects

(b) Ratings of LR stacks averaged by subject



(c) Ratings of HR reconstructed volume from Niftymic



(d) Ratings of HR reconstructed volume from NeSVoR

Figure 6: Histograms of ratings of LR input stacks (a), averaged by subject (b), and HR reconstructed volume slices from Niftymic (c) and NeSVoR (d) algorithms averaged over the two *runs*.

Histogram (a) shows sparsity in the input LR stacks quality. Yet, histogram (b) displays a more compact distribution of ratings when the stack ratings are averaged per subject. From this observation, even though one of its input stack displays poor quality, the other stacks might be still be overall of satisfying quality for the reconstruction to perform well.

Histogram (c) and (d) displays the HR reconstructed volume ratings distribution from Niftymic and NeSVoR respectively. One can see that Niftymic reconstruction ratings are very heterogeneous, with 10 unreadable subject reconstructions along with 10 of relatively higher quality. Histogram (d) also presents for NeSVoR 60% of the subject reconstructions with a poor rating, below 1.5. These observations should be taken into account in the discussions for the uncertainty metric performance analysis.

3.2 Results

3.2.1 Outlier Rejection Study

Niftymic Outlier Rejector In this section, the outlier rejection technique of Niftymic algorithm is analysed for further processing. Figure 7 presents Niftymic algorithm process of the outliers after the reconstructions. One can observe the imbalanced representation of outliers on the extremas of the stack. Indeed, more than half of the slices were rejected. The Figure 8 (a) demonstrates the robustness added to Niftymic algorithm with the outlier rejection.

In average, 78% of the slices located in the middle of the stack are kept for the final reconstruction. This also means that Niftymic outlier rejection functions discard outlier slices that should have contained useful information.. Figure 8 (b) presents an example of Niftymic rejecting falsely a slice from the final reconstruction.

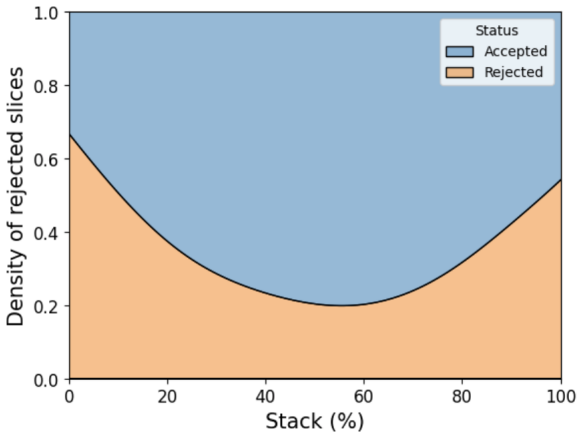
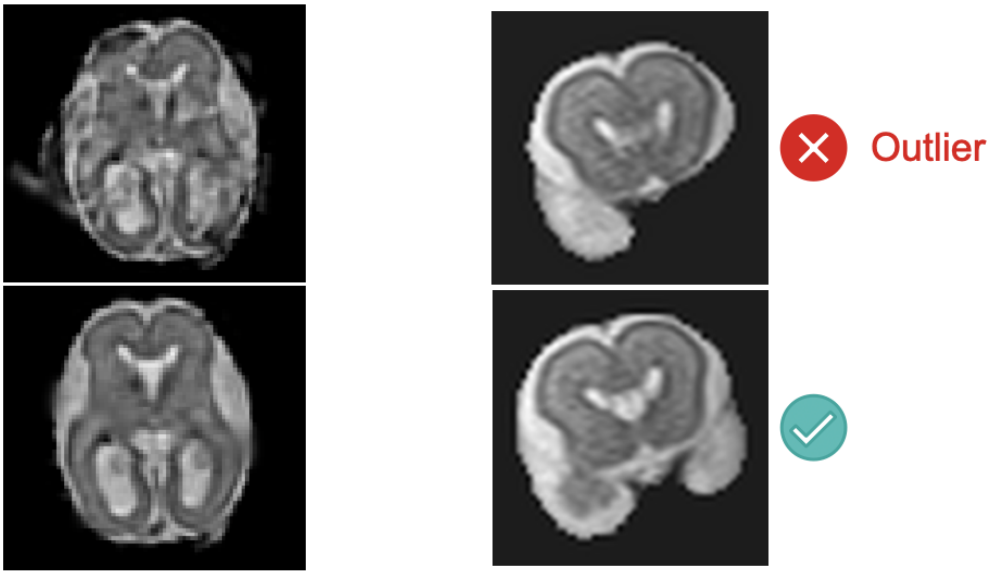


Figure 7: Outlier rejection of input slices from Niftymic reconstruction. The rejected slices are mostly located on the periphery of the stacks, where less information of the fetal brain is available. The results was averaged over all input stacks from all studied subjects.

Ultimately, even-though Niftymic outlier rejection is necessary for the final reconstruction quality, there are slices wrongly rejected which could participate in decreasing the accuracy of the reconstruction. If a small part of a specific slice contain disruptive artefacts it would be appropriate to remove this specific part. This technique is approached by NeSVoR algorithm.



(a) (b)

Figure 8: (a): Top image is a slice in the axial view of the HR reconstructed volume using Niftymic algorithm without outlier rejection. Bottom image represents the same slice reconstructed using outlier rejection. (b) Example input LR slices in coronal view, brain mask applied, of rejected slice (above) and accepted slice (below) for Niftymic reconstruction on a subject.

NeSVoR Outlier Rejector Pixel outliers approach from NeSVoR is an adequate solution to keep the maximum amount of useful information from the input slices. This technique allows precise localisation in the input stacks of the artefacts. Indeed, Figure 9 displays an opposite characteristic to the one presented for Niftymic in Figure 7. It demonstrates that the pixels outliers, with high σ^2 values, are not necessarily located on the periphery of the volume, which was the case for the slice outliers in stacks in Niftymic reconstructions. The assumption is made that the middle slice of a stack should approximately represent a central area of the brain. Therefore, NeSVoR outlier rejection technique perform the outlier rejection uniformly across the stacks. It illustrates an ability to localize

smaller disruptive areas rather than complete slices.

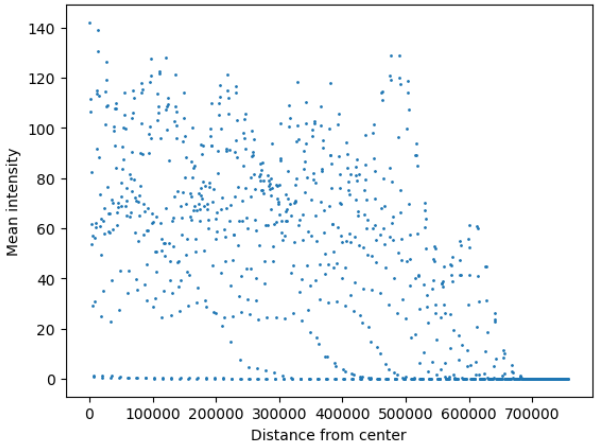


Figure 9: Voxel mean $\sigma^2(x)$ intensity of HR volume by NeSVoR algorithm with respect to the relative distance to the center of the reconstructed fetal brain volume. The mean intensities were computed on $N_{bins} = 1000$ distance bin ranges $\mathcal{D}_i = [\min(D) + id_{bin}, \min(D) + (i+1)d_{bin}]$ where $d_{bin} = \frac{\max(D)-\min(D)}{N_{bins}}$, $i \in \{0, N_{bins} - 1\}$, and D is the complete array of voxel relative distances to the center of the volume. This ensures, results are not biased by the imbalanced representation of pixels far from the center.

These two outlier rejection methods are trained to detect artefacts, and discard inputs that are corrupted too much. Then, one could expect that these methods holds information on the corruption of the slices. In the following section we experiment the different metrics built out of outlier rejection parameters from each algorithm. This metrics aim at displaying the corruption information of the input data.

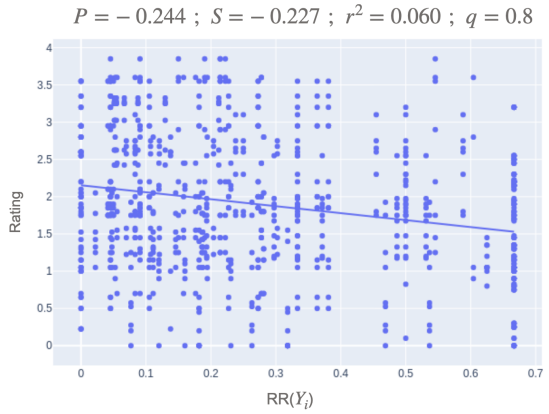
3.2.2 Low-resolution quality control experiments

The LR ratings are used to evaluate a potential correlation with the uncertainty metrics $\mathrm{AR}(Y_i)$, $\mathrm{RR}(Y_i)$ for Niftymic. For NeSVoR, 22 metric variables were tested for the experiment. The main metrics for both algorithm are defined in part 2.4. Recalling this part, $\mathrm{RR}(Y_i)$ and the main metrics for NeSVoR should correlate negatively with the results while $\mathrm{AR}(Y_i)$ should correlate positively. We also recall that metrics for NeSVoR can be normalized per subject.

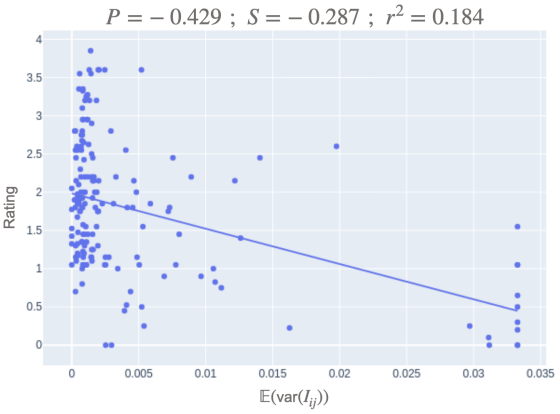
The scatter plots in Figure 10 presents the correlation results for both algorithms. For the two algorithms, only the metric which present the highest correlation with the results is displayed. In scatter plot (a), this metric is $\mathrm{RR}(x)$, which still displays a very low correlation with the ratings. The correlation is negative, which was expected, but the r^2 value is too low for assuming any dependence between the two variables.

Scatter plot (c) also present a very low correlation with the ratings. The normalized metrics for NeSVoR are not predictive of the ratings in any way. Yet, the absolute metric $\mathbb{E}(\mathrm{var}(I_{ij}))$ in scatter plot (b) displays a slightly higher correlation with ratings. The data is not normally distributed so it violates the hypothesis of Pearson correlation. One can still observe a relatively high Spearman correlation $S = -0.287$ with a non-negligeable $r^2 = 0.194$ value. The correlation is negative as expected.

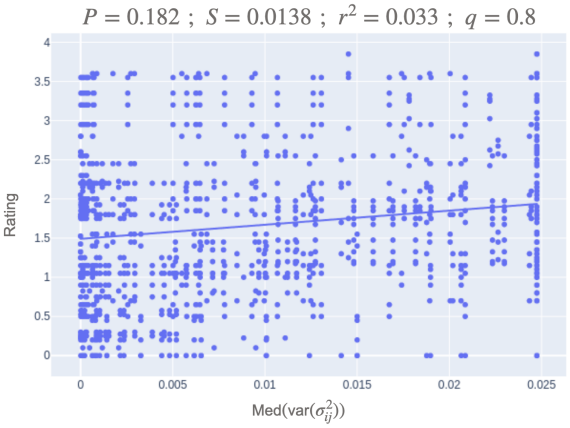
Using outlier rejection parameters from each algorithms, different metrics were tested for predicting the LR input stacks quality. Yet, the outlier rejection parameters can also be used to evaluate the performance of the algorithm itself. If the algorithm reject a greater amount of information, one could observe a lower quality in the reconstructed volume if the outlier rejection is not perfectly built. The following section presents the evaluation of the metrics for HR volume uncertainty evaluations. We should then correlate these metrics with the reconstructed volume ratings to evaluate their performance as QC metrics.



(a) Niftymic: Relative Ratio

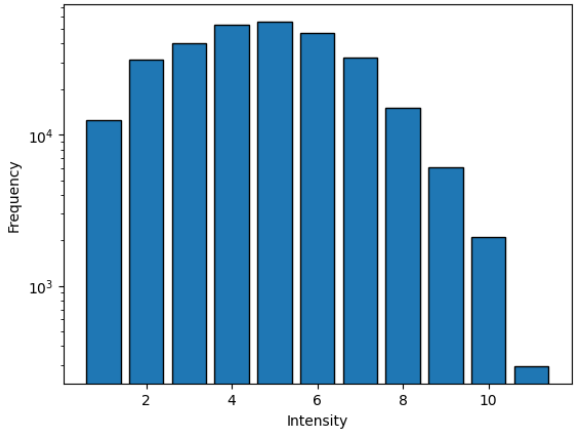


(b) NeSVoR: Mean variance of absolute intensities

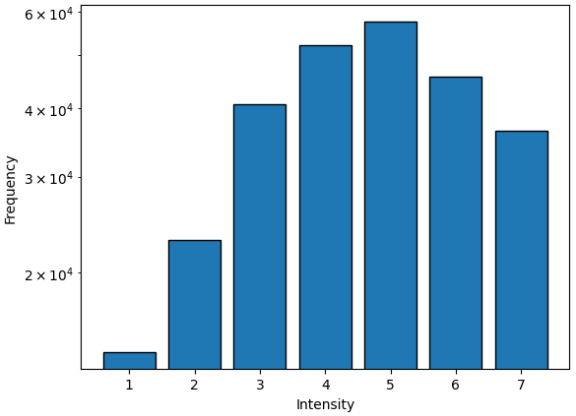


(c) NeSVoR: Median variance of relative intensities

Figure 10: Scatter plot of different uncertainty measurements to the corresponding rating of low-resolution stacks. Pearson P, Spearman S correlation values are displayed with the r^2 value. Some metrics were clipped to a quantile for which the value q is displayed. The linear regression of the points is also represented. The metrics displayed demonstrated the highest value of correlation across all tested metrics



(a) Absolute confidence



(b) Relative uncertainty

Figure 11: Voxel intensity histograms of the two confidence/uncertainty metrics from a Niftymic reconstructed HR volume.

3.2.3 Uncertainty Maps

Niftymic uncertainty maps The computation of the uncertainty maps is added inside Niftymic algorithm pipeline so that the metrics are directly obtained along with the HR reconstructed volume. To start with, Figure 11 displays the distribution of metrics value. One can already observe that AC and RU metrics present very similar results. It is expected from their definition in Eq. 14, 15. An example of the two metrics as confidence/uncertainty maps on the same subject is presented on Figure 12.

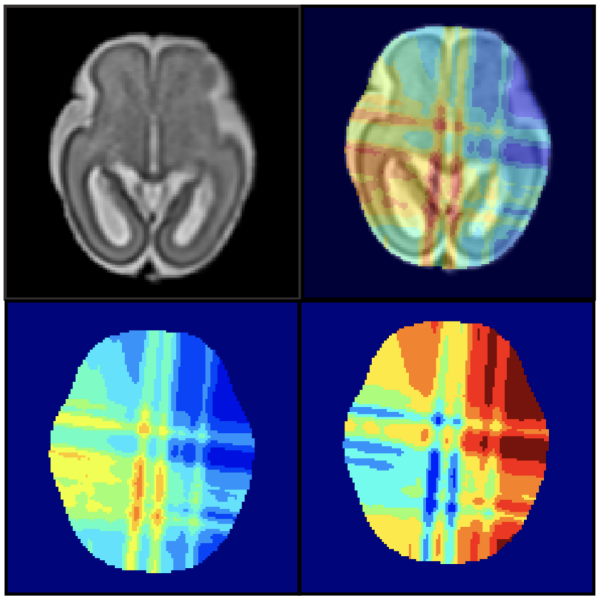


Figure 12: HR reconstructed volume form Niftymic slice along with corresponding uncertainty metric maps. Top left, a slice from the HR reconstructed volume in the axial view of the fetal brain. Top right, the slice overlayed with the absolute confidence AC(x) map. Bottom left, the absolute confidence AC(x) map. Bottom right, the relative uncertainty map RC(x). Same subject as for Figure 11

RU is supposed to display more contrasts as it is rescaled relatively to the subject uncertainty values. These sharper contrasts are quite noticeable on the middle left comparing AC and RU maps. RU maps presents higher number of segmentation areas in that specific area.

Secondly, as the metric values are integers, the segmentation in different confidence/uncertainty levels is easily performed. From close observation of the segmented areas, one can distinct line/rectangular shaped areas. This is due to the outlier rejection and the removal of slices.

Lastly, the top right part of the HR volume slice present indistinct/blurry edges compared to the left part. Comparing with AC and RU measurements, the two metrics predict a lower confidence and higher uncertainty respectively. This blurriness is then probably due to a higher amount of rejection at that specific area.

NeSVoR uncertainty maps The σ_{ij}^2 values are automatically extracted from the NeSVoR reconstruction. The interpolation to $\sigma^2(x)$ the values in 3D space is performed separately, after the reconstruction for convenience. As stated before these $\sigma^2(x)$ values represent a potential metric uncertainty evaluation. The intensity histogram of the $\sigma^2(x)$ values is displayed on Figure 13. One can recognize an exponential decrease of the high-intensity $\sigma^2(x)$ values. The logarithmic term in Eq. 5 is the one leading the loss function as σ^2 values increase. As the loss function penalizes high-intensity σ^2 values with this logarithmic function, the frequency of appearance naturally follows the exponential decrease. The exponential distribution of the σ^2 values is representative of an imbalanced repartition that should be corrected for evaluating a consistent uncertainty scale from this parameter.

An example of the $\sigma^2(x)$ map is displayed on Figure 14. Colour gradient was readjusted logarithmically to ease the observation of the different regions. Indeed, as stated before, the exponential distribution of the σ intensities complicate the reading of the different uncertainty regions. Comparing with Niftymic results, the high uncertainty areas are localised around small areas, pixel-wise. This

makes visual comparison with the HR reconstructed volume slice harder. Regions with high σ^2 values describe input stack pixels that participated less in the final reconstruction. If the high-intensity area is small, it is harder to visualize the possible resulting blurriness in the final reconstruction. This was not the case for Niftymic algorithm.

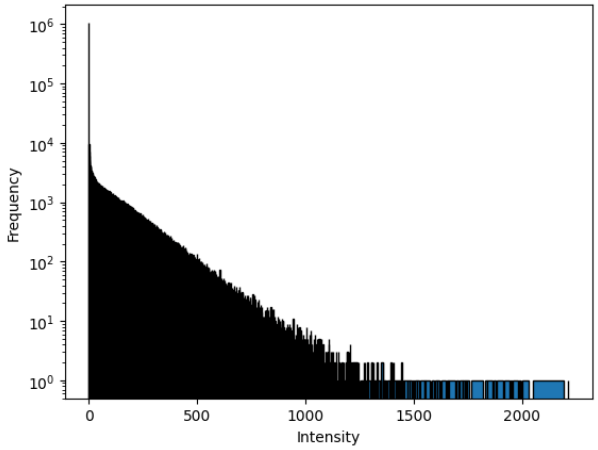


Figure 13: Voxel intensity histograms of the $\sigma^2(x)$ values extracted after the reconstruction performed by NeSVoR. Same subject as the one studied in Figures 12

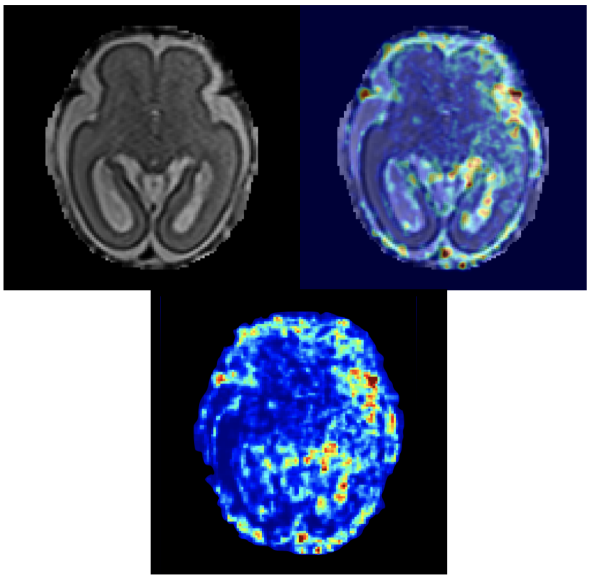


Figure 14: Comparison of the NeSVoR algorithm output and the uncertainty maps. Top left, a slice from the HR reconstructed volume. Top right, the slice overlayed with the $\sigma^2(x)$ value maps. Bottom is the $\sigma^2(x)$ value map. Same subject as the one studied in Figure 12

Ultimately, these metrics have shown potential for evaluating the uncertainty of the HR reconstructions. The following section presents an evaluation of the performance of these metrics using HR reconstruction quality ratings.

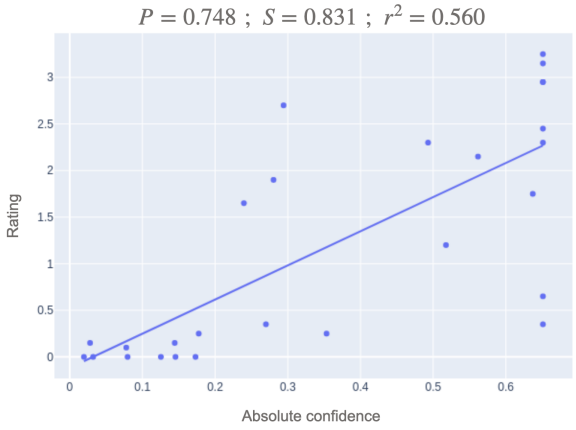
3.2.4 Super-resolution quality control experiments

Different uncertainty metrics for each algorithms are put under performance evaluation. These metrics were mainly presented in part 2.4. To recall part 2.4, AC(x) and the main metrics for NeSVoR are expected to have a positive correlation with the quality ratings. The metrics have a single value at each voxel in the reconstructed volume. For the correlation, the metric values are averaged over the

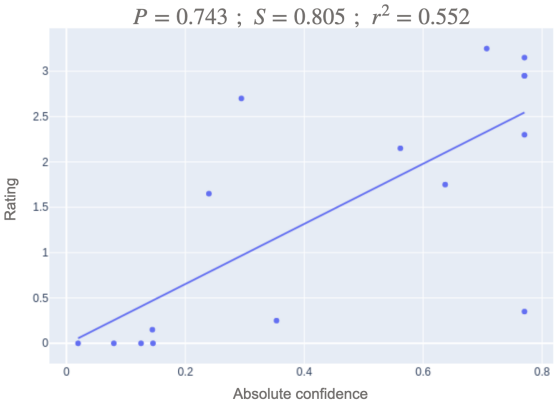
whole volume space Ω. The HR reconstruction quality ratings are used as ground-truth data to test if the metrics are capable of correctly estimating the uncertainty in the final reconstruction.

For each algorithm, there are two sets of initial stacks available as stated in part 3.1. The results are presented as follows: for each algorithm, a scatter plot for each run and both runs combined are produced. The scatter plot displays the uncertainty metric with the highest correlation results with the ratings.

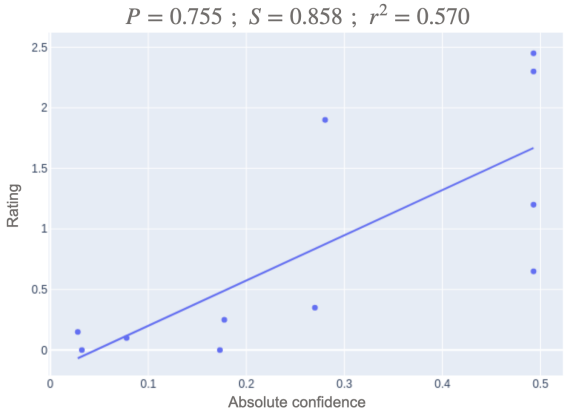
The results for Niftymic algorithm are represented on Figure 15. The correlations are above 0.8 for the three different runs. AC is the most predictive uncertainty metric of the two metrics built for Niftymic algorithm. *run-1* metrics evaluations apparently present a slightly higher correlation with the ratings than for the *run-0*. The ratings for *run-1* only ranges from 0 to 2.5 which characterizes low quality ratings.



(a) Niftymic: $\mathbb{E}(AC(x))$



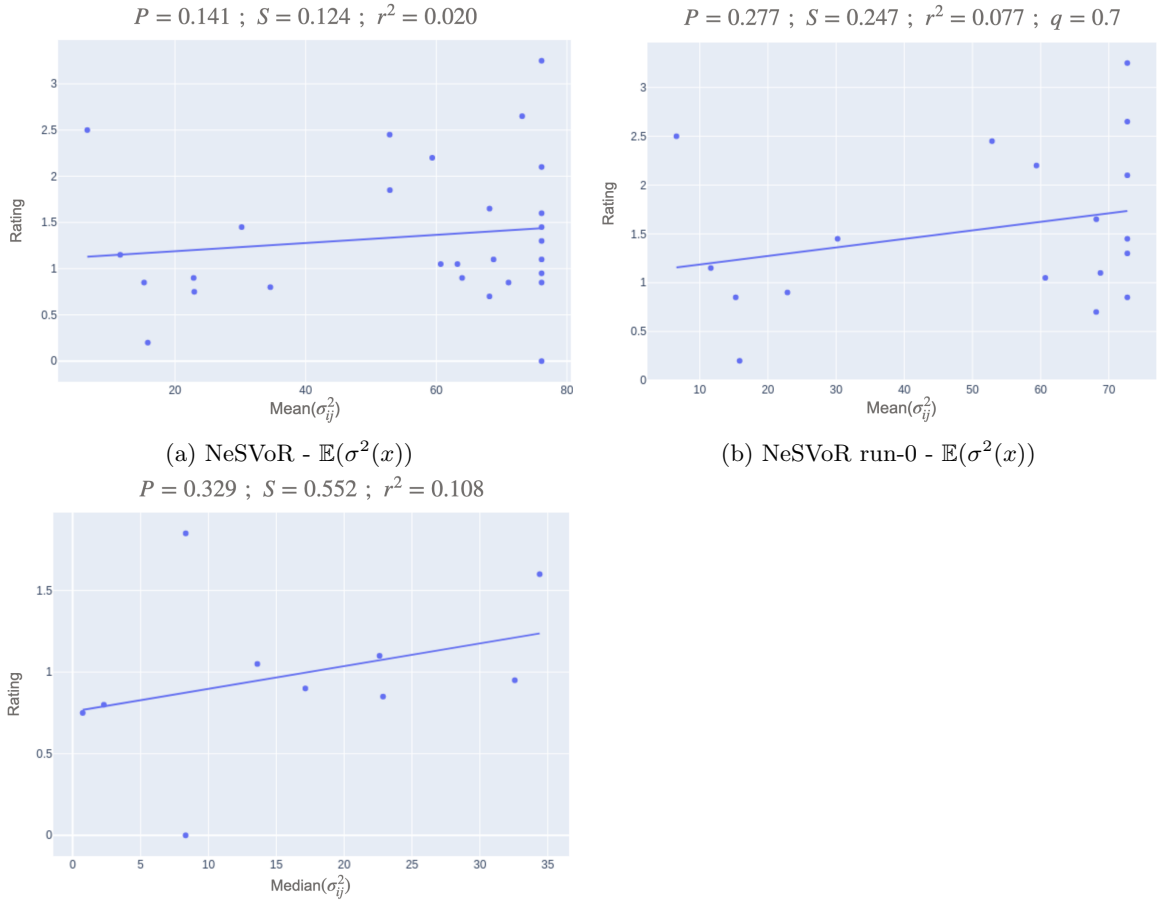
(b) Niftymic run-0: $\mathbb{E}(AC(x))$



(c) Niftymic run-1: $\mathbb{E}(AC(x))$

Figure 15: Scatter plot of different uncertainty measurements to the corresponding rating of high-resolution volume reconstructed by Niftymic. Pearson P, Spearman S correlation values are displayed with the r^2 value. The linear regression of the points is also represented. The metrics displayed demonstrated the highest value of correlation across all tested metrics

The correlation results for NeSVoR algorithm are presented on Figure 16. The highest correlations for the combined runs, the *run-0* and the *run-1* reconstructions are respectively with $\mathbb{E}(\sigma^2(x))$, $\mathbb{E}(\sigma^2(x))$ and Median$(\sigma^2(x))$ metrics. For the combined runs and *run-0* reconstructions, the correlation value is meaningless with a very low $r^2 < 0.1$. However, *run-1* reconstructions display a remarkable Spearman correlation $S = 0.552$. One can also observe that *run-1* reconstructions have their ratings in the range between 0 and 2 which characterizes poor quality ratings.



(a) NeSVoR - $\mathbb{E}(\sigma^2(x))$

(b) NeSVoR run-0 - $\mathbb{E}(\sigma^2(x))$

(c) NeSVoR run-1 - $\mathbb{E}(\sigma^2(x))$

Figure 16: Scatter plot of different uncertainty measurements to the corresponding rating of high-resolution volume reconstructed by NeSVoR. Pearson P, Spearman S correlation values are displayed with the r^2 value. Some metrics were clipped to a quantile for which the value q is displayed. The linear regression of the points is also represented. The metrics displayed demonstrated the highest value of correlation across all tested metrics

4 Discussions

4.1 Significance of Metrics for Quality Control: Reliability, Complementarity, and Limitations

Low-resolution quality predictor metrics For both algorithms, the LR metrics were not able to adequately correlate with the LR input stack ratings. For Niftymic scatter plot in Figure 10 (a), one should focus on the data points located on the upper right and bottom left parts of the plot, . Indeed, initially the metric $\mathrm{RR}(Y_i)$ was expected to correlate negatively with the ratings. The data points from these parts of the plot are the ones responsible for not having a significantly negative slope. For the upper-right part, some stacks are rated very high but contain a high percentage of rejected slice in them. These stacks to misclassified elements from the outlier rejector. Indeed, more than 50% of the slices were rejected from these stacks. A high rating also indicates that these stacks should have brought useful information for the reconstruction. Similarly, the algorithm can fail to identify outlier slices. The bottom left part of the plot presents many poor stack ratings along with a low rejection rate of the slices in the stacks. These issues from the outlier rejection methods can explain why we could not find a significant negative correlation between the $\mathrm{RR}(x)$ metric and the LR stack ratings.

Even though the scatter plot (c) displays the $\mathrm{Median}(\mathrm{var}(I^2_{ij}))$ metric normalized variable, it also means that $\mathbb{E}(\sigma^2_{ij})$ normalized is not correlated with the LR ratings. For NeSVoR, the metrics are less interpretable. Yet, the same remarks can be made as for Niftymic algorithm. Poor quality stacks are

evaluated with low σ^2 intensities, which would correspond to a low amount of pixel outliers. There might be misclassification of pixel outliers which conducts either in falsely high predicted ratings of the inverse. The absolute metric displayed in scatterplot (b) still displays a significant negative correlation with the LR ratings along with a significant r^2 value. Yet, the bottom right part of the graph displays an outlier group from the rest of the scatter plot points. These bottom right located points show a very high value of $\mathbb{E}(\text{var}(I_{ij}))$ along with a very low rating. Indeed, for high-values of this particular metric, the ratings are always low. The inverse is not necessarily true. A deeper analysis should be conducted for any conclusion, but we can hypothesize that this metric can be of use: if a stack has a high value for this uncertainty metric, it is probably a poor quality stack.

High-resolution quality predictor metrics The estimation of the algorithm performance is closely related to the measurement of the confidence in the final reconstruction. In part 3.2.4, different metrics were measured through the SRR algorithm reconstruction process. Niftymic's confidence metrics $\text{AC}(x)$ displayed strong positive correlation $S = 0.831$ with the HR reconstruction ratings in the case of combined runs (see Figure 15). The sparsity of the data points on the scatter plots suggests that $\text{AC}(x)$ is not necessarily a very accurate predictor. Indeed, the r^2 value, which is representing the fraction of the data points variance explained by an other variable, is only of $r^2 = 0.560$. Although the metric is an approximate predictor of the rating, it should be refined before any clinical use. Additionally, the small size of the dataset obliges restriction on any definitive conclusion. One could notice a small difference in the correlations between *run-0* and *run-1*. The *run-1* reconstructions has lower ratings in general than *run-0*. The metrics apparently predicts more adequately poor quality reconstructions.

For the NeSVoR HR reconstructions, the metrics aimed at predicting the quality of the reconstructed volume. Yet, the metrics are measured, from the algorithm point of view, as a potential uncertainty estimate. The algorithms will naturally try to fit its parameter to reduce the uncertainty we try to measure. But this uncertainty may not converge towards the quality rater requirements for a good rating. This challenge is the reason for the development of many different metrics for this algorithm. The results on Figure 16 display in general poor correlation between the different metrics and the HR reconstruction ratings. For scatter plots (a) and (b), the correlations are meaningless which indicates that the metrics are not able to seize the quality rater requirements. For the combined runs case, reconstructions with high ratings are mostly located on the high $\mathbb{E}(\sigma^2(x))$ values, which should be linked to a higher rejection of pixels. The metric grades high quality reconstructions with high uncertainties. The *run-1* still show a remarkable Spearman correlation for rating ranges between 0 and 2, which corresponds to poor quality reconstructions. The metric is still able to approximately predict the ratings of poor quality reconstructions.

4.2 Functionality of the Uncertainty Quantification maps

The uncertainty maps presented in part 3.2.3 are visual inspections of the different metrics studied in the work. The Figure 12 displayed an example of the map capabilities in predicting high/low accuracy areas in the final reconstructed volume.

However, the uncertainty maps did not endure any performance evaluation since it would require pixel-wise ratings of the high-resolution reconstructed volumes. This work is too time-consuming which makes this kind of data inaccessible. For now, the evaluation of the HR uncertainty maps is done through an average over the complete volume.

5 Conclusion

The Niftymic and NeSVoR algorithms enables Super-Resolution Reconstruction of a 3D isotropic volume of the fetal brain from Low-resolution input stacks of 2D MRI slices. These algorithms are outlier-robust, meaning they would discard the input data not assisting adequately the reconstructions. To recall the initial motivations, the outlier rejection methods used in the SRR algorithms could be predictors of the uncertainty of the fetal brain reconstructions. Indeed, it implies discarding information from the final reconstruction. The hypothesis is that uncertainty metrics can be extracted from these outlier rejection methods. Different parameters of each algorithms were extracted to build different possible candidates for uncertainty measurements. For Niftymic, the metrics mostly rely on the quantity of rejected slices during the reconstruction. For NeSVoR, the aim is to retrieve the parameter σ_{ij}^2 and its related variables, as it represents pixel-wise outlier smooth rejection. The

performance metrics should then be evaluated to assess their prediction capability of the uncertainty. First, certain metrics aimed at predicting the quality ratings of the Low-resolution input stacks. Metrics from both algorithms showed poor correlation and predictive power of the quality ratings of the input stacks. The potential issue comes from the outlier rejection methods. If the methods do not perform perfectly the correlations would surely not be observed. Secondly, different metrics aimed at predicting the uncertainty in the final reconstructed HR volume. The metric AC(x) which simply counts the number of slices participating in each voxel of the reconstructed volume, reached high Spearman correlation of $S = 0.831$ with the quality ratings of the volumes. For NeSVoR algorithm, the different metrics did not display any clear correlation with the quality ratings of the reconstructions.

The uncertainty maps could be evaluated more precisely with slice-wise quality ratings of the HR reconstructed volume. This would refine the analysis of the metrics performance. Such uncertainty maps could help clinicians assess the confidence they could have in the HR reconstructed volume.

References

[1] “”Registration-based approach for reconstruction of high-resolution in utero fetal MR brain images.””. In: *Academic radiology* 13.9 (2006).

[2] Manganaro et al. “Fetal MRI of the central nervous system: State-of-the-art.” In: *European Journal of Radiology* 93 (2017), pp. 273–283.

[3] M. Ebner and et al. “An automated framework for localization, segmentation, and super-resolution reconstruction of fetal brain MRI.” In: *NeuroImage* (2020).

[4] O. Esteban and et al. “MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites.” In: *PloS One* 12 (2017), e0184661.

[5] A. Gholipour, J. A. Estroff, and S. K. Warfield. “Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI.” In: *IEEE Trans Med Imaging* 29 (2010), pp. 1739–1758.

[6] B. Kainz and et al. “Fast volume reconstruction from motion corrupted stacks of 2D slices.” In: *IEEE Trans Med Imaging* 34 (2015), pp. 1550–1564.

[7] M. Kuklisova-Murgasova et al. “Reconstruction of fetal brain MRI with intensity matching and complete outlier removal.” In: *Med Image Anal* 16 (2012), pp. 1550–1564.

[8] D. Levine. “Ultrasound versus magnetic resonance imaging in fetal evaluation.” In: *Top Magn Reson Imaging* 12 (2001), pp. 25–38.

[9] S. N. Saleem. “Fetal MRI: An approach to practice: A review.” In: *J Adv Res* 5 (2014), pp. 507–523.

[10] T. Sanchez et al. “FetMRQC: Automated Quality Control for fetal brain MRI”. In: (2023).

[11] S. Tourbier and et al. “MIAL super-resolution toolkit v2.0.1”. In: *Zenodo* (2020).

[12] A. Uus and et al. “Retrospective motion correction in fetal MRI for clinical applications: existing methods, applications and integration into clinical practice.” In: *Br J Radiol* (2022).

[13] J. Xu and et al. “NeSVoR: Implicit neural representation for slice-to-volume reconstruction in MRI.” In: *IEEE Trans Med Imaging* (2023).