Reduction of the cardiac pulsation artifact and improvement of lesion conspicuity in flow-compensated diffusion images in the liver—A quantitative evaluation of postprocessing algorithms

Tobit Führes1 | Marc Saake1 | Jennifer Lorenz1 | Hannes Seuss1,2 | Alto Stemmer3 | Thomas Benkert3 | Michael Uder1 | Frederik Bernd Laun1

Purpose: To enhance image quality of flow-compensated diffusion-weighted liver MRI data by increasing the lesion conspicuity and reducing the cardiac pulsation artifact using postprocessing algorithms.

Methods: Diffusion-weighted image data of 40 patients with liver lesions had been acquired at 1.5 T. These data were postprocessed with 5 different algorithms (weighted averaging, p-mean, percentile, outlier exclusion, and exception set). Four image properties of the postprocessed data were evaluated for optimizing the algorithm parameters. These properties were the lesion to tissue contrast-to-noise ratio (CNR), the reduction of the cardiac pulsation artifact, the data consistency, and the vessel darkness. They were combined into a total quality score ($Q_{total}$, set to 1 for the trace-weighted reference image), which was used to rate the image quality objectively.

Results: The weighted averaging algorithm performed best according to the total quality score ($Q_{total} = 1.111 \pm 0.067$). The further ranking was outlier exclusion algorithm ($Q_{total} = 1.086 \pm 0.061$), p-mean algorithm ($Q_{total} = 1.045 \pm 0.049$), percentile algorithm ($Q_{total} = 1.012 \pm 0.049$), and exception set algorithm ($Q_{total} = 0.957 \pm 0.027$). All optimized algorithms except for the exception set algorithm corrected the pulsation artifact and increased the lesion CNR. Changes in $Q_{total}$ were significant for all optimized algorithms except for the percentile algorithm. Liver ADC was significantly reduced (except for the exception set algorithm), particularly in the left lobe.

Conclusion: Postprocessing algorithms should be used for flow-compensated liver DWI. The proposed weighted averaging algorithm seems to be suited...
best to increase the image quality of artifact-corrupted flow-compensated diffusion-weighted liver data.

**KEYWORDS**
algorithms, cardiac pulsation artifact, diffusion, flow-compensation, liver, postprocessing

1 | INTRODUCTION

DWI of the liver is a valuable tool to detect liver lesions efficiently.\(^1\)\(^-\)\(^5\) However, one main drawback of this technique is the cardiac pulsation artifact, which can lead to signal dropouts, especially in the left liver lobe.\(^6\)\(^-\)\(^9\) In patient data, this can lead to major problems such as decreased visibility or missed lesions, which might lead to a wrong diagnosis.\(^10\) Proposed methods to mitigate this artifact are electrocardiography (ECG) triggers,\(^11\)\(^,\)\(^12\) flow-compensated (FloCo) diffusion encodings,\(^13\)\(^,\)\(^14\) and postprocessing approaches.\(^15\)\(^,\)\(^16\)

ECG triggering allows one to acquire the images at a defined time point of the cardiac cycle. Because the cardiac pulsation artifact rather occurs during the systole than during the diastole,\(^6\) one could perform data acquisition only at these time points. However, this would lead to longer acquisition times. In addition, the diffusion sequence’s fast switching gradients can influence the ECG signal and lead to erroneous detection of trigger points, making this method unreliable.

FloCo diffusion encodings rephase spins that move at a constant velocity so that they do not experience an erroneous dephasing and therefore, do not cause an undesired signal drop.\(^17\)\(^-\)\(^22\) Indeed, FloCo diffusion encodings have been shown to strongly mitigate the severity of the cardiac pulsation artifact.\(^7\) With the rise of ever stronger gradient systems and optimized implementations,\(^23\)\(^-\)\(^25\) the drawback of a reduced \(b\)-value efficiency\(^26\) of FloCo diffusion encodings becomes negligible if one wants to maintain the echo times currently used in liver DWI.\(^24\)\(^,\)\(^27\) Laun et al\(^27\) recently showed that the use of FloCo diffusion encodings indeed resulted in a better diagnostic performance in assessing focal liver lesions. However, although FloCo diffusion encodings performed well, they could not fully compensate for the cardiac pulsation artifact.

Postprocessing schemes usually aim to detect erroneous signal dropouts and have been shown to improve the image quality.\(^15\)\(^,\)\(^28\) These schemes can fail, however, if too much data are corrupted. This may happen in DWI of the left liver lobe, especially if conventional diffusion encodings are used, for which >50% of source images can be affected by the pulsation artifact.\(^7\)

Therefore, a combination of FloCo diffusion encodings with advanced postprocessing schemes seems to be a promising next step. The FloCo diffusion encoding would be the first line of defense against the pulsation artifact. The encoding would increase data quality to such an extent that advanced postprocessing schemes could more easily, reliably, and robustly handle the compensation of the remaining pulsation artifact.

One problem with postprocessing approaches is that one usually must make trade-offs between image features improved by the postprocessing scheme and those that are worsened. For example, taking a pixel-wise maximum of repeatedly acquired images minimizes pulsation-induced signal drops and can increase the visibility of lesions. However, it also amplifies the influence of artifactually high signals such as that of residual blood signal in DWI (which was reported to be of particular relevance in FloCo DWI of the liver\(^7\)). Therefore, rating only favorable features such as the lesion to liver contrast-to-noise ratio (CNR) would disregard the importance of unfavorable features like the appearance of these bright blood signals. This emphasizes that multiple image features should be considered to evaluate algorithms holistically and objectively.

Therefore, this study aimed to find an optimal postprocessing scheme for diagnosing of focal liver lesions in FloCo DWI. Five postprocessing algorithms were considered, and the algorithms’ parameters were optimized with respect to mitigation of the cardiac pulsation artifact, CNR of lesions, the darkness of vessels, and the general data consistency using an automated evaluation pipeline. This automated evaluation was used to be able to optimize even complex algorithms with multiple free parameters efficiently.

2 | METHODS

2.1 | Acquisition

The data were acquired in a prospective study that had been approved by the local Institutional Ethics Committee (study number 276_19 B). Each participant provided written informed consent. The same source data had been used in a previous evaluation, focusing on comparing FloCo and conventional diffusion encodings, without consideration of postprocessing algorithms.\(^27\) Between January and August 2020, 40 patients (age range, 34-74 years, mean
2.2 | Algorithms

Only the $b = 800 \text{s/mm}^2$ images were considered because they are known to be more greatly affected by the pulsation artifact than $b = 50 \text{s/mm}^2$ images. The signal for each voxel was calculated from a set of 12 signals from the 12 source images. For all evaluations, the magnitude images were used.

As a reference image, a trace-weighted image was computed using the vendor-specific calculation approach (Figure 1A). First, the arithmetic average of the 4 images per diffusion direction was computed, and second, the geometric average of the resulting images was computed:

$$S = \sqrt[4]{\prod_{d=x,y,z} \frac{1}{4} \sum_{m=1}^{4} S_{dm}} = \sqrt[4]{S_{x1} + S_{y1} + S_{z1} + S_{x1}}.$$  

For example, $S_{x1}$ denotes the signal of the first image of diffusion direction $x$. This calculation was done for each voxel separately. Unlike this notation suggests, the actual diffusion encoding directions were not parallel to the scanner’s $x$-, $y$-, and $z$-axis but to the above-defined diffusion directions.

Five correction algorithms were implemented in Python (version 3.7.9). To ensure the reference image and the results of the evaluated algorithms were comparable, the geometric averaging was maintained. The source code of the algorithms is available on GitHub (https://github.com/20tf22/postprocessing_floc).

2.2.1 | Algorithm 1: Weighted averaging

Weighted averaging (Figure 1B). Gaussian-filtered versions of the original images were used as weight maps $w_{dm}$, where the index $d$ denotes the diffusion direction and $m$ the repetition.

$$w_{dm} = \text{Gauss}_{k_{s1} \times k_{s1}}(x, y, z) \cdot S_{dm},$$

where the star denotes the discrete convolution operator in position space (i.e., in the image plane) and $k_{s1}$ is the size of the filter kernel. The standard deviation $\sigma_1$ of the Gaussian function was calculated from the kernel size $k_{s1}$ as follows: $\sigma_1 = 0.3 \times ((k_{s1} - 1) \times 0.5 - 1) + 0.8$, where the definition of the Python Library OpenCV (version 4.5.5), was used.

The weight was thus higher for image regions of higher signal intensity, reflecting that the pulsation artifact usually results in signal drops. The reason for the application of the Gaussian filter was to suppress the effect of single or few voxel outliers on the used weights, which might arise, e.g., from blood in small vessels. Because the regions affected by the pulsation artifact are usually much larger than such outlier regions, the weights were thought to still capture the presence of the pulsation artifacts well.

Inspired by Ichikawa et al, the final image was calculated as follows:

$$S_{\text{weighted averaging}} = \sqrt[4]{\prod_{d=x,y,z} \sum_{m=1}^{4} \frac{w_{dm}^n S_{dm}}{\sum_{m=1}^{4} w_{dm}^n}}.$$  

$n$ is the exponent of the weight maps and increases the weighting of bright regions. The kernel size $k_{s1}$ was sampled in steps of 2 voxels ($= 3.125 \text{ mm}$) from 1 voxel (no filter applied) to 49 voxels ($= 7.7 \text{ cm}$), $n$ was sampled in steps of 1 from 1 to 10.

2.2.2 | Algorithm 2: $p$-mean algorithm

$p$-Mean algorithm (similarly proposed by Liau et al) (Figure 1C). The resulting image was calculated as follows:

$$S_{p\text{-mean}} = \sqrt[4]{\sum_{m=1}^{4} \left( \frac{S_{x1} + S_{y1} + S_{z1} + S_{x1}}{4} \right)^p}.$$  

The parameter $p$ can assume any positive value. This calculation is performed for each voxel separately. $p$ was sampled in steps of 0.1 between $p = 1$ and $p = 20$. 
FIGURE 1  Visualization of reference image calculation and postprocessing algorithms. “Dir.” means “diffusion direction”, the green arrows denote (weighted) arithmetic averaging, and the blue arrows denote (weighted) geometric averaging. Optimized parameters are printed in red. (A) Reference image calculation. (B) Weighted averaging. (C) $p$-Mean. (D) Percentile. (E) Outlier exclusion. (F) Exception set. More information on the described algorithms are provided in the main text.
2.2.3 Algorithm 3: Percentile algorithm

Percentile algorithm (Figure 1D). The resulting image was calculated as the q-th percentile (with linear interpolation between the 2 closest signals if the percentile did not exactly match a signal value) of the single images:

\[ S_{\text{new}} = \text{perc}_q(S_{x1}, S_{x2}, S_{x3}, S_{x4}, S_{y1}, S_{y2}, S_{y3}, S_{y4}, S_c, S_{s2}, S_{s2}, S_{s3}, S_{y4}) \]

This calculation was done for each voxel separately. q was sampled in steps of 5 from 0 to 100. The 0th, 50th, and 100th percentiles are equivalent to minimum, median, and maximum, respectively.

2.2.4 Algorithm 4: Outlier exclusion algorithm

Outlier exclusion algorithm (Figure 1E). This iterative approach uses the fact that pulsation artifacts induce a signal dropout rather than a signal increase. The basic concept is inspired by the “informed Restore” algorithm.28

Step 1. A deviation map \( S_{\text{dev}} \) is calculated, which represents the relative standard deviation of all 12 images:

\[ S_{\text{dev}} = \text{std}(S_{x1}, S_{x2}, S_{x3}, S_{x4}, S_{y1}, S_{y2}, S_{y3}, S_{y4}, S_c, S_{s2}, S_{s2}, S_{s3}, S_{y4}) / S. \]

\( S \) is the reference image, and “std” is the standard deviation.

Step 2. The deviation map \( S_{\text{dev}} \) was smoothed using a Gaussian filter.

\[ S_{\text{dev.smoothed}} = \text{Gauss}_{k_s} \times k_s \ast S_{\text{dev}}. \]

Again, the intention of the filtering was to minimize the effect of small outlier regions. For the definition of the Gaussian function, see above (Algorithm 1).  

Step 3. (a) Only regions with \( S_{\text{dev.smoothed}} > \text{thr} \) were considered for the exclusion of data points. \( \text{thr} \) is a value between 0 and 1. For each voxel in these regions, the lowest of the signal values was determined and excluded as an outlier. Note that the excluded values may originate from different images for adjacent pixels. (b, first iteration only) For each voxel with \( S_{\text{dev.smoothed}} > \text{thr} \), \( \text{std}_{\text{no max}} = \text{std} \) (all signals except the highest signal) and \( \text{std}_{\text{no min}} = \text{std} \) (all signals except the lowest signal) were calculated and smoothed (Gaussian filter, kernel size empirically set to \((7 \text{ px})^2 = (1.1 \text{ cm})^2\)). If \( \text{std}_{\text{no max}} < 0.8 \text{ std}_{\text{no min}} \), this meant that the deviation of the highest signal dominated the standard deviation. Then, this high signal was excluded additionally. This step aimed to prevent the algorithm from continuously excluding low values if they were not the reason for the high \( S_{\text{dev.smoothed}} \) value.

Step 4. A new image \( S_{\text{new}} \) was calculated using the remaining signal values with a weighted geometric mean. For example, if \( S_{y2} \) was removed:

\[
S_{\text{new}} = \sqrt[4]{\left(\frac{S_1 + S_5 + S_4}{3}\right) \cdot \left(\frac{S_1 + S_5 + S_4}{3}\right) \cdot \left(\frac{S_5 + S_4 + S_2}{4}\right) \cdot \left(\frac{S_1 + S_5 + S_4}{3}\right)},
\]

where the original signals (i.e., those without application of the Gaussian filter) were used.

Step 5. A new relative standard deviation was calculated:

\[ S_{\text{dev}} = \text{std}(\text{remaining signals}) / S_{\text{new}}. \]

where the original signals were used (i.e., those without a Gaussian filter applied).

The process continues at Step 2. The Steps 2-5 were repeated \( k \) times. \( k \) is also called “number of iterations” in the following text. When the algorithm had finished, the current \( S_{\text{new}} \) was used as result.

\( k_{s2} \) was sampled in steps of 2 voxels (= 3.125 mm) from 1 voxel (no filtering) to 79 voxels (= 12.3 cm), \( \text{thr} \) in steps of 0.05 from 0.1 to 0.7, and \( k \) from 3 to 10.

2.2.5 Algorithm 5: Exception set algorithm

Exception set algorithm (by Arning et al)30 (Figure 1F). This algorithm finds the subset \( S_j \) (called “exception set”) of the 12 elements \( S \), whose exclusion maximizes the homogeneity of the remaining set while not excluding more elements than necessary.

\[ \sum_{i=0}^{\text{maxout}} \binom{12}{i} \] possible signal subsets \( S_j \) with up to \( \text{maxout} \) elements are calculated. For each subset, a smoothing factor \( \text{SF} \) is calculated:

\[ \text{SF} (S_j) = |S \setminus S_j| \cdot (\text{Var}(S) - \text{Var}(S \setminus S_j)) \]

where \(| \cdot |\) denotes the number of elements in the set and \(\text{Var}(\cdot)\) denotes the variance of the elements in the set. Therefore, \(\text{SF}\) is higher for exception sets that are small and whose exclusion decreases the variance of the remaining signals.

For each voxel, the set \( S_j \) with the highest \(\text{SF} \) was excluded. The new image \( S_{\text{new}} \) was calculated with a weighted geometric mean (as described in Algorithm 4, Step 4). \( \text{maxout} \) was sampled from 1 to 12.

2.3 Optimization and evaluation

In the reference images of 29 of 40 patients, pulsation artifacts could be observed by visual inspection. Therefore, data from these patients were used for the evaluation.

For each patient, eight 2D regions of interest were segmented on the 8800 images. The segmentations were performed with the software MITK (version v2021.2), by a
physicist with 3 years of experience in abdominal imaging (T.F.), who was supervised by a board-certified radiologist (M.S.) with 10 years experience in abdominal imaging. The eight segmented 2D regions were:

Region 1 (R1). The region in the left liver lobe affected by the pulsation artifact.

Region 2 (R2). A 10 cm² circular region in the right lobe of apparently healthy liver tissue that was visually unaffected by the pulsation artifact, sparing major vessels and lesions.

Region 3 (R3). Visible major vessels.

Region 4 (R4). A circular region with a size of 40 cm² in the right liver lobe, intentionally not sparing vessels and lesions.

Region 5 (R5). A lesion in the left liver lobe, which was affected by the pulsation artifact (if possible).

Region 6 (R6). The surrounding, apparently healthy tissue of lesion (R5).

Region 7 (R7). A lesion in the right liver lobe, which was not affected by the pulsation artifact (if possible).

Region 8 (R8). The surrounding, apparently healthy tissue of lesion (R7).

Note that not all regions were necessarily located in the same slice. R1–R4 were segmented in all slices affected by the pulsation artifact. The lesions and their corresponding surrounding tissues were segmented in the same slices (which might differ from those with R1–R4). Representative segmentations are shown in Figure 2.

Using these segmentations, 4 subscores were calculated. The subscores were used to quantify the enhancement in image quality. All subscores were designed such that a desirable effect (such as good data consistency or no pulsation artifact) resulted in a high value.

1. Pulsation artifact subscore $\overline{PA}$,
   $$\overline{PA} = \frac{\text{mean}(R2)}{\text{mean}(R1)}.$$
   This subscore quantifies the signal ratio between a region that is affected by the pulsation artifact and a region that is not.

2. Vessel darkness subscore $\overline{VD}$,
   $$\overline{VD} = \frac{\text{mean}(R2) - \text{mean}(R3)}{\text{mean}(R2)}.$$
   This subscore quantifies the relative signal difference between vessels and tissue. Dark vessels lead to a high value.

3. Data consistency subscore $\overline{DC}$,
   $$\overline{DC} = 1 - \text{mean} \left( \frac{|R4 - R4(\text{reference image})|}{R4(\text{reference image})} \right).$$
   This subscore quantifies the deviation between the image calculated by the algorithm and the original image. The subtraction from 1 ensures that the score is high for low deviations.

4. CNR subscores for the lesions in the left and the right lobe,
   $$\overline{CNR}_{left} = \frac{\text{mean}(R5) - \text{mean}(R6)}{\text{std}(R6)}, \quad \overline{CNR}_{right} = \frac{\text{mean}(R7) - \text{mean}(R8)}{\text{std}(R8)}.$$

For every slice containing segmentations, the subscores were calculated for the reference image and for the images obtained with every possible combination of algorithm parameters.

The subscores were normalized to the value they yielded for the reference image (i.e., $PA = \overline{PA}/\overline{PA}$ (reference image), and analogously for the other subscores).

The subscores were averaged over all slices, which means that exactly 1 set of subscores was obtained for each combination of algorithm parameters. Additionally, the standard deviation over all slices was calculated for each combination of algorithm parameters.

From the subscores, a total quality score $Q_{total}$ was calculated:

$$Q_{total} = \frac{(PA + VD + DC + CNR)}{4},$$

where $CNR = \frac{\overline{CNR}_{left} + \overline{CNR}_{right}}{2}$. Because the subscores were calculated per slice and the lesions were not necessarily located in the same slice, it was not possible to average both CNR values earlier. The standard deviations of the CNR subscore and of $Q_{total}$ were calculated using error propagation.

For the reference image, the total quality score is 1. A higher value indicates an increase in image quality. The parameter set, which yielded the highest $Q_{total}$ score, was chosen as optimized parameter set.

To test whether the total quality score differs significantly from 1, a 2-sided 1-sample $t$ test was used. The significance level was set to 0.05. Statistical analyses were performed with the Python modules SciPy\textsuperscript{31} (version 1.7.3) and statsmodels\textsuperscript{32} (version 0.13.2).

### 2.4 ADC calculation

ADC maps were calculated with the original images (using the original $b50$ and original $b800$ image) and with the edited images (using the original $b50$ and the edited $b800$ images) of all 5 optimized algorithms. The algorithms were not applied to the $b50$ images because they are less affected by the pulsation artifact\textsuperscript{15} and because only 1 image instead of 4 images per diffusion direction was recorded. The mean ADC was determined in the left liver lobe (R1), the right liver lobe (R2), a lesion in the left liver lobe (R5), and a lesion in the right liver lobe (R7) and subsequently averaged over all patients. Statistical analysis (of the unaveraged values) was done with a dependent $t$ test for paired samples. The significance level was set to 0.05.
3 | RESULTS

3.1 | Weighted averaging

Figure 3A shows $Q_{\text{total}}$ in dependence of the kernel size $k_s$ and exponent $n$. The highest total quality score of 1.111 was achieved for a kernel size $k_s = 23$ voxels ($= 3.6 \text{ cm}$) and an exponent $n$ of 4. For these values, the image quality according to $Q_{\text{total}}$ changed significantly ($p < 0.0001$). Along the $n$-direction, a well-defined maximum occurred for all $k_s$. Without filtering the image ($k_s = 1$), the maximum occurred at $n = 2$, and it shifted to $n = 4$ for larger $k_s$. Along the $k_s$ direction, the maximum occurred around $k_s = 20$ for low values of $n$ and shifted to higher values for larger values of $n$. In the 2 further plots of Figure 3A, 1 parameter is fixed to these optimal values (i.e., to $k_s = 23$ voxels or $n = 4$), and the dependence on the respective other parameter is shown. Additionally, the subscores are shown. Again, with increasing $n$ and $k_s$, the total quality score reached a maximum at $n = 4$ and $k_s = 23$, respectively, and decreased at larger values. In Figures 3B,C, representative images, edited with different parameters of $k_s$ and $n$, are shown. The window/level settings of all images, which are compared directly in this article, are identical. All parameter combinations brightened the left liver lobe and increased the conspicuity of lesions. With increasing $k_s$, the images appear smoother and the vessels are more clearly visible. For high $n$ (e.g., $n = 10$), the visibility of vessels decreased.

In summary, the total quality score sets the following demand on the free parameter $p$:

• approximately $3 \leq p \leq 6$, best $p \approx 4.5$.

3.2 | $p$-Mean algorithm

In Figure 4, the different subscores and the total quality score $Q_{\text{total}}$ are plotted against $p$, which is the only free parameter of the algorithm. With increasing $p$, the vessel darkness (VD) and data consistency (DC) scores decreased, whereas the pulsation artifact (PA) score increased. CNR reached a maximum and then remained relatively constant. The highest $Q_{\text{total}}$ of 1.045 was achieved for $p = 4.5$. For this value, the image quality according to $Q_{\text{total}}$ changed significantly ($p = 0.0044$). Additionally, representative images are shown. With rising $p$, the left liver lobe and vessels appear brighter, and the images appear noisier.

These results may be interpreted as follows. The total quality score sets the following demand on the free parameter $p$:

• approximately $3 \leq p \leq 6$, best $p \approx 4.5$.

3.3 | Percentile algorithm

Figure 5 shows the results of the percentile algorithm for different values of the percentile $q$. With increasing $q$, the total quality score $Q_{\text{total}}$ increased, then reached the highest value of 1.012 at $q = 75$ and decreased again. The image quality according to $Q_{\text{total}}$ did not change significantly ($p = 0.4760$) The representative images in the figure show that for low values of $q$, the left liver lobe and lesions are hardly visible. With rising $q$, the image becomes brighter, the pulsation artifact decreases, and the lesions become more clearly visible. For very high values of $q$, however, the image noise increases strongly, leading to decreased vessel and lesion visibility.

In summary, the total quality score sets the following demands on the parameter $q$ of the percentile algorithm: approximately $70 < q < 80$.

3.4 | Outlier exclusion algorithm

Figure 6 shows the results of the outlier exclusion algorithm. Because this algorithm has 3 free parameters, the total quality score is a 3D scalar field, which is difficult to
FIGURE 3 Weighted averaging algorithm. (A, left) Total quality score in dependence of exponent \( n \) and kernel size \( k_{s_1} \). In the central and right plot, the right axis refers to the total quality score and is scaled differently. (A, center) Total quality score (filled points) and sub-scores (crosses) for \( n = 4 \) in dependence of \( k_{s_1} \). (A, right) Total quality score and sub-scores for \( k_{s_1} = 23 \) voxels (= 3.6 cm) as a function of \( n \). To avoid overlapping error bars, the errors of the sub-scores are shown as colored regions. PA, pulsation artifact subscore; VD, vessel darkness subscore; DC, data consistency subscore; CNR, contrast-to-noise ratio. (B) The reference image and images edited with different kernel sizes. With a higher kernel size, the image appears smoother and vessels are more clearly visible. The arrowheads mark a poorly-visible vessel; the arrow marks a lesion with increased conspicuity. (C) Images edited with different powers of the weights. Higher \( n \) increases the brightness of the left liver lobe, very high \( n \) decreases vessel visibility. The arrowheads mark a poorly visible vessel.

visualize. For this reason, the total quality score in dependence of 2 of 3 free parameters is shown in Figure 6A. The third free parameter is fixed to the value at which \( Q_{\text{total}} \) becomes maximal. The highest total quality score of 1.086 was achieved for the kernel size \( k_{s_2} = 41 \) voxels (= 6.4 cm), the correction threshold \( thr = 0.3 \), and the number of iterations \( k = 8 \). For these values, the image quality according to \( Q_{\text{total}} \) changed significantly (\( p < 0.0001 \)). The total quality score showed a pronounced maximum at approximately \( thr = 0.25 \) to 0.45 for all depicted \( k \) and \( k_{s_2} \). Along the \( k \)-dimension, \( Q_{\text{total}} \) increased monotonically until a maximum at \( k = 8 \). For larger \( k \), \( Q_{\text{total}} \) decreased slightly. Along the \( k_{s_2} \)-dimension, \( Q_{\text{total}} \) increased for small values of \( k_{s_2} \), then reached a maximum and decreased again.

In Figure 6B, the dependences of the total quality score and the single scores on a single parameter are shown, with the other 2 parameters fixed to their optimal values. In general, the score increases and saturates with the number of iterations. As the correction threshold and kernel size
increased, the score reached a maximum and decreased again.

Figure 6C (full figure, Supporting Information Figure S1C-E) shows representative images, with 2 parameters fixed and 1 varied. With an increasing number of iterations, the left liver lobe appears brighter (Supporting Information Figure S1C). Increasing the filter kernel size $k_{s2}$ reduces the image noise and leads to a smoother and less artifactual appearance. A false-positive lesion disappears (shown by the thick white arrow), but the increased kernel size decreases the lesion conspicuity (Figure S1D). Low threshold values increase the left liver lobe brightness and lead to relatively high signals and bright vessels. Increasing the threshold to the optimal value restores the original appearance of the right liver lobe while maintaining the pulsation artifact reduction (Figure 6C/ Figure S1E).

In summary, the total quality score sets the following demands on the free parameters of the outlier exclusion algorithm:

- $\text{thr} \approx 0.3$,
- $k \geq 7$, which corresponds to 7 of 12 (i.e., $\sim$60% of the available data), and
- $k_{s2} \sim 30-50$ voxels, which corresponds to 5-8 cm.

### 3.5 Exception set algorithm

Figure 7 shows the total quality score and the subscores of the exception set algorithm. The total quality score was lower than 1, indicating a worsening image quality. The highest score of 0.957 was achieved for $\text{maxout} = 1$. The score decreased until $\text{maxout} \approx 6$ and then stayed approximately constant. The image quality according to $Q_{\text{total}}$ changed significantly ($p < 0.0001$). The shown representative image indicates a reduction of lesion visibility for all $\text{maxout}$. Even at $\text{maxout} = 1$, the image appears slightly noisier, and artifactual regions appear (e.g., around the spleen). With rising $\text{maxout}$, the visual impression...
of the image becomes worse. The image appears noisier, the left liver lobe looks darker than in the reference image, and the lesion in the right liver lobe is no longer visible.

In summary, $Q_{\text{total}}$ and the visual impression indicate that the exception set algorithm should not be used for postprocessing FloCo DWI liver image data.

### 3.6 Comparing the optimized algorithms

The total quality scores and the subscores obtained with the respective optimized parameter set are summarized in Table 1. Four of 5 analyzed algorithms increase the image quality according to the total quality score $Q_{\text{total}}$. The weighted averaging algorithm features the highest $Q_{\text{total}}$ (= 1.111), followed by outlier exclusion (= 1.086), $p$-mean (= 1.045), and percentile (= 1.012). The exception set algorithm does not increase the image quality ($Q_{\text{total}}$ = 0.957).

Weighted averaging performs rather well in all subcategories (rank 2, 2, 1, and 3 in the subscores PA, VD, CNR, and DC, respectively). Outlier exclusion has the highest subscore in 2 categories (PA and DC) and exception set in 1 category (VD).

In Figure 8 (Supporting Information Figure S2), representative images are shown. All algorithms except for the exception set algorithm increase the lesion conspicuity and correct the cardiac pulsation artifact to a certain degree. The visual appearance of the optimized outlier exclusion algorithm, weighted averaging, and $p$-mean are similar. Nonetheless, the lesion marked by an arrow in the third row is best visible with weighted averaging. The images appear to be the most detailed with weighted averaging (second row, marked by arrowheads). The percentile algorithm induces noise in the images. The exception set algorithm darkens the image and decreases lesion conspicuity.

### 3.7 ADC evaluation

The results of the ADC calculations are shown in Figure 9 and, together with the respective $P$-values, in Supporting Information Table S1. Representative ADC maps are shown in Supporting Information Figure S3. In general, both ADC values calculated in the left lobe strongly decreased for all algorithms except the exception set algorithm. In the right liver lobe, both ADC values decreased rather slightly. The exception set algorithm nearly induced no change.
FIGURE 6  Outlier exclusion algorithm. (A) Total quality score with 1 free parameter fixed at its optimal value and the dependence on the other 2 parameters is shown. (B) Total quality score (filled points) and subscores (crosses). In each plot, 2 free parameters are fixed at their optimal values and the dependence on the other parameter is shown. The right axis refers to the total quality score and is scaled differently. In the center plot (dependence on $k_2$), only every second data point is shown to increase clarity. To avoid overlapping error bars, the errors of the subscores are shown as colored regions. CNR, contrast-to-noise ratio; DC, data consistency subscore; PA, pulsation artifact subscore, VD, vessel darkness subscore. (C) Reference image and images, edited with different values of the parameter $\text{thr}$. The increase of $\text{thr}$ from a low value to the optimal value restores the original appearance of the right liver lobe while maintaining the pulsation artifact reduction. The black arrowheads mark a poorly visible vessel. The full figure showing representative images for all parameter dependencies can be found in Supporting Information Figure S1.

4 | DISCUSSION

In this work, 5 postprocessing algorithms were systematically investigated with respect to their ability to minimize the pulsation artifact in FloCo liver DWI and to increase lesion CNR while maintaining data consistency and vessel darkness. Four of 5 algorithms increased the image quality according to the defined total quality score $Q_{\text{total}}$. This score quantitatively evaluated the image quality and consisted of 4 subscores (pulsation artifact reduction, data consistency, lesion CNR, and retention of dark vessels).

The definition of the subscores was aimed to correlate well with visual impressions. For example, an increase in the PA score of 20% should approximately correlate with a visual impression of a 20% pulsation artifact reduction. Therefore, we tended to define the subscores as simply as mathematically possible. Of course, other subscore definitions would have led to somewhat different optimized parameter sets. However, we think that the benefit of an
FIGURE 7 Exception set algorithm. (Top right) Total quality score (filled points) and subscores (crosses) for the exception set algorithm. The right axis refers to the score and is scaled differently. To avoid overlapping error bars, the errors of the subscores are shown as colored regions. (Surrounding images) The algorithm does not increase the image quality. With rising values for maxout, the image gets noisy and artifactual. The arrow in the reference image marks a lesion that is no longer visible for \( maxout \geq 5 \).

TABLE 1 Values of the subscores and the total quality score \( Q_{\text{total}} \) obtained with the respective optimized parameter set

|               | PA subscore | VD subscore | CNR subscore | DC subscore | \( Q_{\text{total}} \) | \( P\)-value * |
|---------------|-------------|-------------|--------------|-------------|----------------|----------------|
|               | Left        | Right       |              |             |                |                |
| Weighted averaging | 1.362 ± 0.150 | 0.894 ± 0.069 | 1.263 ± 0.10 | 0.926 ± 0.034 | 1.111 ± 0.067 | <0.0001        |
|                |             |             | 1.317 ± 0.360 |              |                |                |
|                |             |             | 1.209 ± 0.217 |              |                |                |
| \( p\)-mean   | 1.235 ± 0.088 | 0.793 ± 0.075 | 1.263 ± 0.154 | 0.790 ± 0.038 | 1.045 ± 0.049 | 0.0044         |
|                |             |             | 1.296 ± 0.225 |              |                |                |
|                |             |             | 1.230 ± 0.211 |              |                |                |
| Percentile    | 1.263 ± 0.093 | 0.801 ± 0.091 | 1.185 ± 0.162 | 0.797 ± 0.043 | 1.012 ± 0.049 | 0.4760         |
|                |             |             | 1.181 ± 0.239 |              |                |                |
|                |             |             | 1.190 ± 0.220 |              |                |                |
| Outlier exclusion | 1.439 ± 0.153 | 0.826 ± 0.121 | 1.139 ± 0.136 | 0.940 ± 0.054 | 1.086 ± 0.061 | <0.0001        |
|                |             |             | 1.225 ± 0.252 |              |                |                |
|                |             |             | 1.054 ± 0.100 |              |                |                |
| Exception set | 0.942 ± 0.047 | 1.075 ± 0.031 | 0.866 ± 0.091 | 0.946 ± 0.011 | 0.957 ± 0.027 | <0.0001        |
|                |             |             | 0.827 ± 0.144 |              |                |                |
|                |             |             | 0.904 ± 0.109 |              |                |                |

Abbreviations: PA, pulsation artifact; VD, vessel darkness; DC, data consistency; CNR, contrast-to-noise ratio.

*\( P\)-values refer to the null hypothesis that \( Q_{\text{total}} \) is equal to 1.
FIGURE 8  Reference and optimized images of 3 patients. The white rectangles in the overview images mark the areas, which are shown below. The visual impression of images processed with weighted averaging, p-mean, and outlier exclusion algorithm is similar. The percentile algorithm induces noise in the images. (First column) A small lesion in the left liver lobe is hardly visible in the reference image (see arrow). All shown algorithms substantially increase lesion conspicuity. (Second column) A strong pulsation artifact is visible in the reference image. All shown algorithms substantially reduce the artifact. The weighted averaging algorithm seems to reproduce the original anatomy best (e.g., small horizontal tissue structure, see arrowheads). (Third column) The lesion conspicuity increases with all shown algorithms. The arrow marks a small lesion that is most clearly visible with the weighted averaging algorithm. The full figure with the comparison of all algorithms can be found in Supporting Information Figure S2
objective, unbiased image quality evaluation and the possibility to optimize the algorithm parameters automatically is higher than the potential dependence of the parameter set on the subscore definition. The same can be stated for the subscore weightings. We decided to weigh them all equally to have clear and straightforward dependencies. Here again, other weightings might change the exact positions of the found optima, but there was no conclusive argument for this, and we assumed that the overall picture would not have changed. In the case of changed requirements (e.g., the lower weighting of lesion CNR in healthy volunteer measurements), this could be easily implemented by weighting the subscores in the calculation of $Q_{\text{total}}$ differently.

According to the $Q_{\text{total}}$ score, weighted averaging increased the image quality most. The visual impression of the respective images in Figure 8 are images with well-visible lesions and strongly reduced pulsation artifact. With relatively small kernel sizes for blurring the weight maps, it would be possible to increase the CNR and pulsation artifact scores slightly. However, this can only be achieved by a lower data consistency and brighter vessels. With the kernel size used for the optimized value, the data consistency is close to 1, and the appearance of the vessels can be kept dark. Ichikawa et al., who also evaluated a weighted averaging approach, did not use blurring of the weight maps (in terms of our study, kernel size one). They stated that artifacts with high signals posed a problem for their approach. However, this is not a problem if weight map blurring is introduced, at least for artifacts of limited size, because the blurring diminishes the influence of single artifactually bright voxels.

The $p$-mean algorithm had the best performance for $p = 4.5$. This is in very good agreement with the study by Liau et al., who first proposed this method and found that $p = 4$ was suited best. However, the $p$-mean algorithm only leads to a limited enhancement of the overall image quality because the vessel darkness could not be retained sufficiently and because the algorithm changed the image impression considerably (low data consistency). The visual impression was one of a slightly noisy image as previously reported by Liau et al. One reason for this is that we were unable to implement a suitable low-pass filtering that could have resulted in a smoother image.

Although easy to implement and understand, the percentile algorithm led to only a slight increase in the image quality according to $Q_{\text{total}}$. The limited quality enhancement may be because pulsation artifact correction and vessel darkness are 2 competing goals. When dark areas of the left liver lobe were brightened, the dark vessels were brightened too. In other algorithms (like weighted averaging), this was prevented by low-pass filtering, meaning that only larger dark areas are corrected. In particular, median ($q = 50$) and maximum ($q = 100$) operations do not appear well suited for image correction. The median does not lead to sufficient pulsation artifact correction, and the maximum leads to a noisy image with poorly visible vessels.

The outlier exclusion algorithm can substantially enhance the image quality according to the $Q_{\text{total}}$ score. Our implementation was similar to that of the “informed Restore” algorithm, variants of which have been applied successfully in several settings and been implemented in publicly available libraries. In our study, the algorithm also performed quite well. The possibility to exclude the highest value once (Step 3b in the Methods section) is crucial when a “wrong” maximum in the 12 single pixel values is present. For example, this can happen because of respiratory motion. Because the images were acquired under free breathing conditions, sometimes 1 of 12 images (3 diffusion directions × 4 repetitions) showed an adjacent slice. Bright lesions from this wrong slice were the reason for
the high standard deviation and led to the iterative exclusion of low values. In the final image, the lesions were then visible although they were located in the adjacent slice. The optimization also showed that it was better to allow more values to be excluded and use a relatively low threshold for exclusion rather than a high threshold with only a few allowed iterations. Apparently, the high threshold prevented many areas from being corrected by the algorithm. This is confirmed by the threshold value of 0.7, where all scores are ~1, as in the reference image.

The exception set algorithm was not well suited for postprocessing our image data, although it appeared to be a somewhat sophisticated approach. The main reason is that in the corrupted data, often >50% of the samples were of poor quality, although FloCo diffusion encodings were used. With the exception set approach, outlier detection can only work when the majority of the samples are of sufficient data quality. Here, the algorithm tends to exclude the high values, because they are the minority and seen as outliers, while the low values are maintained. This results in even greater decreased visibility of already poorly visible lesions and an increased pulsation artifact.

The changes of the ADC values in the left lobe can be explained by the changed brightness of the edited b8000 images in this region. The algorithms that performed well in terms of this study decreased the ADC, whereas the exception set algorithm induced an increase. In the right lobe, there are still significant changes, although they are much smaller. The reason is presumably that the pulsation artifact is also present in the right lobe in some cases, albeit to a much reduced extent. The outlier exclusion algorithm performs better in this case, because it only acts on regions below a certain quality threshold.

This is also a first limitation to this study; some algorithms might change quantitative image parameters of uncorrupted regions. Second, the acquisition in free breathing, which was the only possibility with the prototypical sequence used, led to small shifts between averages of the same slice, which might have produced some wrong signal outliers from other slices. Third, the resulting images were not rated by experienced radiologists. However, an in-depth optimization of the algorithms by experienced readers would have required evaluation of $40 \times 13 \times 8$ (kernel sizes $\times$ threshold values $\times$ iterations) = 4160 parameter combinations for the outlier exclusion algorithm alone. Therefore, this can only be seen as a second possible step, in which radiologists rate the already optimized algorithms and is planned for further studies. Fourth, only data from a single scanner of a single vendor were used. Therefore, our findings might be of limited generalizability. Fifth, we had not implemented a dedicated quality assurance protocol (i.e., with polyvinylpyrrolidone phantoms). In conclusion, postprocessing algorithms should be used for liver DWI. Out of the considered algorithms, the weighted averaging algorithm seems to be suited best to increase the image quality of FloCo liver diffusion images.

ACKNOWLEDGMENTS
Funding by the Deutsche Forschungsgemeinschaft is gratefully acknowledged (446875476 and 430650228). Open Access funding enabled and organized by Projekt DEAL. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST
The co-authors Alto Stemmer and Thomas Benkert are employed at Siemens Healthcare.

DATA AVAILABILITY STATEMENT
The code that supports the findings of this study is openly available in GitHub at https://github.com/20tf22/postprocessing_floco.

ORCID
Tobit Führes https://orcid.org/0000-0001-6804-6353

REFERENCES
1. Ayuso C, Rimola J, Vilana R, et al. Diagnosis and staging of hepatocellular carcinoma (HCC): current guidelines. Eur J Radiol. 2018;101:72-81.
2. Donato H, Franca M, Candelaria I, Caseiro-Alves F. Liver MRI: from basic protocol to advanced techniques. Eur J Radiol. 2017;93:30-39.
3. Koh DM, Collins DJ. Diffusion-weighted MRI in the body: applications and challenges in oncology. AJR Am J Roentgenol. 2007;188:1622-1635.
4. Parikh T, Drew SJ, Lee VS, et al. Focal liver lesion detection and characterization with diffusion-weighted MR imaging: comparison with standard breath-hold T2-weighted imaging. Radiology. 2008;246:812-822.
5. Qayyum A. Diffusion-weighted imaging in the abdomen and pelvis: concepts and applications. Radiographics. 2009;29:1797-1810.
6. Kwee TC, Takahara T, Niwa T, et al. Influence of cardiac motion on diffusion-weighted magnetic resonance imaging of the liver. Magn. 2009;22:319-325.
7. Rauh SS, Riehinger AJ, Ohlmer S, et al. A mixed waveform protocol for reduction of the cardiac motion artifact in black-blood diffusion-weighted imaging of the liver. Magn Reson Imaging. 2020;67:59-68.
8. Ozaki M, Inoue Y, Miyati T, et al. Motion artifact reduction of diffusion-weighted MRI of the liver: use of velocity-compensated diffusion gradients combined with tetrahedral gradients. J Magn Reson Imaging. 2013;37:172-178.
9. Riehinger A, Laun FB, Bickelhaupt S, et al. On the dependence of the cardiac motion artifact on the breathing cycle in liver diffusion-weighted imaging. PLoS One. 2020;15:e0239743.
10. Nasu K, Kuroki Y, Nawano S, et al. Hepatic metastases: diffusion-weighted sensitivity-encoding versus SPIO-enhanced MR imaging. *Radiology*. 2006;239:122-130.

11. Xiang Z, Ai Z, Liang J, Li G, Zhu X, Yan X. Evaluation of regional variability and measurement reproducibility of intravoxel incoherent motion diffusion weighted imaging using a cardiac stationary phase based ECG trigger method. *Biomed Res Int.* 2018;2018:4604218-4604211.

12. Metens T, Absil J, Denolin V, Bali MA, Matos C. Liver apparent diffusion coefficient repeatability with individually predetermined optimal cardiac timing and artifact elimination by signal filtering. *J Magn Reson Imaging*. 2016;64:1100-1110.

13. Johnson GA, Maki JH. In vivo measurement of proton diffusion in the presence of coherent motion. *Invest Radiol*. 1991;26:540-545.

14. Wetscherek A, Stieltjes B, Laun FB. Flow-compensated intravoxel incoherent motion diffusion imaging. *Magn Reson Med*. 2015;74:410-419.

15. Liau J, Lee J, Schroeder ME, Sirlin CB, Bydder M. Cardiac motion in diffusion-weighted MRI of the liver: artifact and a method of correction. *J Magn Reson Imaging*. 2012;35:318-327.

16. Chang LC, Walker L, Pierpaoli C. Informed RESTORE: a method for robust estimation of diffusion tensor from low redundancy datasets in the presence of physiological noise artifacts. *Magn Reson Med*. 2012;68:1654-1663.

17. Maki JH, Macfall JR, Johnson GA. The use of gradient flow compensation to separate diffusion and microcirculatory flow in MRI. *Magn Reson Med*. 1991;17:95-107.

18. Fujita N, Harada K, Sakurai K, Akai Y, Kozuka T. Separation of diffusion and slow flow effects by use of flow rephasing and dephasing. *Magn Reson Med*. 1992;24:109-122.

19. Ahlgren A, Knutsson L, Wirestam R, et al. Quantification of microcirculatory parameters by joint analysis of flow-compensated and non-flow-compensated intravoxel incoherent motion (IVIM) data. *NMR Biomed*. 2016;29:640-649.

20. Brockstedt S, Thomsen C, Wirestam R, et al. Use of an enhanced gradient system for diffusion MR imaging with motion-artifact reduction. *Acta Radiol*. 1995;36:662-670.

21. McTavish S, Van AT, Peeters JM, et al. Gradient nonlinearity correction in liver DWI using motion-compensated diffusion encoding waveforms. *MAGMA*. 2021. https://doi.org/10.1007/s10334-021-00981-6.

22. Szczepankiewicz F, Sjolund J, Dall’Armellina E, et al. Motion-compensated gradient waveforms for tensor-valued diffusion encoding by constrained numerical optimization. *Magn Reson Med*. 2021;85:2117-2126.

23. Aliotta E, Wu HH, Ennis DB. Convex optimized diffusion encoding (CODE) gradient waveforms for minimum echo time and bulk motion-compensated diffusion-weighted MRI. *Magn Reson Med*. 2017;77:717-729.

24. Middione MJ, Loecher M, Moulin K, Ennis DB. Optimization methods for magnetic resonance imaging gradient waveform design. *NMR Biomed*. 2020;33:e4308.

25. Zhang Y, Pena-Nogales O, Holmes JH, Hernando D. Motion-robust and blood-suppressed M1-optimized diffusion MR imaging of the liver. *Magn Reson Med*. 2019;82:302-311.

26. Grebenkov DS. NMR survey of reflected Brownian motion. *Rev Mod Phys*. 2007;79:1077-1137.

27. Laun FB, Führes T, Seuss H, et al. Flow-compensated diffusion encoding in MRI for improved liver metastasis detection. *PLoS One*. 2022;17:e0268843.

28. Ichikawa S, Motosugi U, Tamada D, et al. Improving the quality of diffusion-weighted imaging of the left hepatic lobe using weighted averaging of signals from multiple excitations. *Magn Reson Med Sci*. 2019;18:225-232.

29. Bradski G. The openCV library. *Dr. Dobb’s j. softw. tools*. 2000;25:120-123.

30. Arning A, Agrawal R, Raghavan P. A Linear Method for Deviation Detection in Large Databases. Presented at: KDD; 1996.

31. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods*. 2020;17:261-272.

32. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with python. Presented at: *Proceedings of the 9th Python in Science Conference*; 2010.

33. Andersson JLR, Graham MS, Zsoldos E, Sotiropoulos SN. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *Neuroimage*. 2016;141:556-572.

34. Fieremans E, Lemberskiy G, Veraart J, Sigmund ES, Gyftopoulos S, Novikov DS. In vivo measurement of membrane permeability and myofiber size in human muscle using time-dependent diffusion tensor imaging and the random permeable barrier model. *NMR Biomed*. 2017;30:e3612.

35. Pannek K, Raffelt D, Bell C, Mathias JL, Rose SE. HOMOR: higher order model outlier rejection for high b-value MR diffusion data. *Neuroimage*. 2012;63:835-842.

36. Collier Q, Veraart J, Jeurissen B, den Dekker AJ, Sijbers J. Iterative reweighted linear least squares for accurate, fast, and robust estimation of diffusion magnetic resonance parameters. *Magn Reson Med*. 2015;73:2174-2184.

37. Garyfallidis E, Brett M, Amirbekian B, et al. Dipy, a library for the analysis of diffusion MRI data. *Front Neuroinform*. 2014;8:8.

38. Saake M, Seuss H, Riezinger A, et al. Image quality and detection of small focal liver lesions in diffusion-weighted imaging: comparison of navigator tracking and free-breathing acquisition. *Invest Radiol*. 2021;56:579-590.

39. Wagner F, Laun FB, Kuder TA, et al. Temperature and concentration calibration of aqueous polyvinylpyrrolidone (PVP) solutions for isotropic diffusion MRI phantoms. *PLoS One*. 2017;12:e0179276.

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher’s website.

**Figure S1.** Outlier exclusion algorithm. (A) Total quality score with 1 free parameter is fixed at its optimal value and the dependence on the other 2 parameters is shown. (B) Total quality score (filled points) and subscores (crosses). In each plot, 2 free parameters are fixed at their optimal values and the dependence on the other parameter is shown. The right axis refers to the total quality.
score and is scaled differently. In the center plot (dependence on $k_{S^2}$), only every second data point is shown to increase clarity. To avoid overlapping error bars, the errors of the subscores are shown as colored regions. CNR, contrast-to-noise ratio; DC, data consistency subscore; PA, pulsation artifact subscore, VD, vessel darkness subscore.

(C–E) Reference image and images edited with different parameters of the algorithm. Increasing the number of iterations increases the brightness of the left liver lobe. An increase of the kernel size $k_{S^2}$ leads to a smoother and less artifactual image. White arrows mark artifactual image parts. The increase of $thr$ from a low value to the optimal value restores the original appearance of the right liver lobe while maintaining the pulsation artifact reduction. The black arrowheads mark a poorly visible vessel.

**Figure S2.** Reference and optimized images of 3 patients. The white rectangles in the overview images mark the areas, which are shown below. The visual impression of images processed with weighted averaging, $p$-mean, and outlier exclusion algorithm is similar. The percentile algorithm induces noise in the images, and the exception set algorithm darkens the image. (First column) A small lesion in the left liver lobe is hardly visible in the reference image (see arrow). All algorithms except for the exception set algorithm substantially increase lesion conspicuity. (Second column) A strong pulsation artifact is visible in the reference image. All algorithms substantially reduce the artifact except for the exception set algorithm. The weighted averaging algorithm seems to reproduce the original anatomy best (e.g., small horizontal tissue structure, see arrowheads). (Third column) The lesion conspicuity increases with all algorithms except for the exception set algorithm. The arrow marks a small lesion that is most clearly visible with the weighted averaging algorithm.

**Figure S3.** ADC maps calculated with the original images and the edited images.

**Table S1.** ADC values in $\mu m^2/\text{ms}$ in the left liver lobe, the right liver lobe, a lesion in the left lobe, and a lesion in the right lobe, respectively, calculated with the original image and the edited images. Values are averaged over evaluated patients. $P$-values refer to the null hypothesis that the mean of the original ADC and of the respective “new” ADC are equal. These data are visualized in Figure 9.