



Quantification of metal artifacts in computed tomography: methodological considerations

Nils Große Hokamp^{1,2,3}, Brendan Eck², Florian Siedek¹, Daniel Pinto dos Santos¹, Jasmin A. Holz¹, David Maintz¹, Stefan Haneder¹

¹Institute for Diagnostic and Interventional Radiology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany; ²Department of Radiology, Case Western Reserve University, Cleveland, OH, USA; ³Department of Radiology, University Hospitals Medical Center, Cleveland, OH, USA

Correspondence to: Priv.-Doz. Dr. Nils Große Hokamp. Institute for Diagnostic and Interventional Radiology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany; Department of Radiology, Case Western Reserve University, Cleveland, OH, USA; Department of Radiology, University Hospitals Medical Center, Cleveland, OH, USA. Email: nils.grosse-hokamp@uk-koeln.de.

Abstract: Numerous methods for artifact quantification in computed tomography (CT) imaging have been suggested. This study evaluated their utility with regards to correspondence with visual artifact perception and reproducibility. Two titanium rods (5 and 10 mm) were examined with 25 different scanning- and image-reconstruction parameters resulting in different types and extents of artifacts. Four radiologists evaluated every image against each other using an in-house developed software. Rating was repeated two times (2,400 comparisons = 2 times × 4 readers × 300 comparisons). Rankings were combined to obtain a reference ranking. Proposed approaches for artifact quantification include manual measurement of attenuation, standard deviation and noise and sophisticated algorithm-based approaches within the image- and frequency-domain. Two radiologists conducted manual measurements twice while the aforementioned algorithms were implemented within the Matlab-Environment allowing for automated image analysis. The reference ranking was compared to all aforementioned methods for artifact quantification to identify suited approaches. Besides visual analysis, Kappa-statistics and intraclass correlation coefficients (ICC) were used. Intra- and Inter-reader agreements of visual artifact perception were excellent (ICC 0.85–0.92). No quantitative method was able to represent the exact ranking of visually perceived artifacts; however, ICC for manual measurements were low (ICC 0.25–0.97). The method that showed best correspondence and reproducibility used a Fourier-transformed linear ROI and lower-end frequency bins. Automated measurements of artifact extent should be preferred over manual measurements as the latter show a limited reproducibility. One method that allows for automated quantification of such artefacts is made available as an electronic supplement.

Keywords: Computed tomography (CT); quantitative image analysis; artifact; image reconstruction

Submitted Oct 17, 2019. Accepted for publication Mar 30, 2020.

doi: 10.21037/qims.2020.04.03

View this article at: <http://dx.doi.org/10.21037/qims.2020.04.03>

Background

Computed tomography (CT) imaging is frequently impaired by artifacts arising from various kinds of metal hardware and multiple approaches to mitigate these artifacts are available (1-5). They either work via modification of image acquisition, dedicated image reconstruction or their

combination (4-6). Numerous studies are available that evaluate a benefit from one or more of the aforementioned approaches. Most of these studies comprise a quantitative and a visual evaluation of images. Here, visual analysis is considered gold standard regarding evaluation of artifact reduction, while methods for objective artifact quantification vary widely.

Methods based on measurement within regions of interest (ROI) are fairly common; authors either measure attenuation within the most pronounced artifacts, measure the standard deviation within the artifact or within a reference tissues affected by artifacts (3,7-14). Other groups have conducted 1-dimensional size measurements of the metal object or the width of the artifact itself (3,11,15). Some groups suggest to use advanced quantitative methods instead, such methods either work within the image or the frequency domain (5,16,17).

For quantitative assessment of image contrast established parameters such as the signal- and/or contrast-to-noise ratio are available and commonly defined and used. They allow for comparison between studies, vendors and platforms. For artifact quantification such commonly accepted methods are missing to date; yet, they are highly desirable.

In this study, we adapted proposed methods for artifact quantification from literature. Besides manual, multi-reader measurements, automated approaches were adapted within the Matlab-Environment. Some methods were developed further in order to enable transferability to other settings. To demonstrate the validity of quantification algorithms we evaluated their correspondence with visual artifact perception. After all, this study aims to propose guidance for metal artifact quantification in CT imaging.

Methods

Reference scans

To obtain reference images for quantitative and visual analysis, two metal rods with 5- and 10-mm diameter were examined in a 15 cm × 15 cm water phantom using a 64-row dual layer CT scanner (IQon Spectral Detector CT, Philips Healthcare, Best, The Netherlands). Metal rods originated from a medical instrument supplier (L. Klein SA, Biel, Switzerland) and comprise materials actually used for medical applications (e.g., spinal fusions, external fixations). Image acquisition and reconstruction varied in order to obtain a series of 25 reference images representing a wide range of artifacts encountered (*Table 1, Figure 1*).

Reference ranking

To obtain a reference ranking, the aforementioned 25 images were compared in a forced-choice pairwise comparison, i.e., every combination of images was reviewed by the readers, who were asked to choose the image with fewer artifacts.

In total, 300 comparisons were performed using $N_{\text{images}} = 25$ source images $\{N_{\text{comparisons}} = 1/2 \times [(N_{\text{images}})^2 - N_{\text{images}}]\}$. To ensure standardized settings and avoid recording errors, a dedicated software in the Matlab-Environment was developed and used. The software displays two images in a random order side-by-side, the interface then waits for the reader to click on the image with fewer artifacts, upon decision the next pair is presented, and the responses are stored automatically. Images were displayed on a designated image viewing station with fixed window settings (Window center 0, width 150).

In total, 4 readers (2 board certified radiologists with >8-year experience in CT imaging and 2 radiologists-in-training with 4 years of experience in CT imaging) performed the reading twice with more than 4 weeks apart. This resulted in a total of 2400 comparisons (2 × 4 readers × 300 comparisons). All readings were combined to obtain a reference ranking of all images ranging from fewest to most artifacts serving as gold standard in this study.

Intrareader-repeatability and interreader-agreement on a per-comparison basis were determined using the intraclass-correlation coefficient (ICC).

Artifact quantification algorithms: Part 1—manual methods

Manual measurements were conducted by the same 4 readers using a clinical image viewer (Agfa ImpaxEE, Bonn, Germany). The readers were trained based on 5 images not included in the final analysis and conducted the reading twice with 4 weeks apart to avoid recall bias. Between these manual measurements and the aforementioned reference ranking analysis there was a time lapse of 6 months. As suggested in earlier studies, the readers measured widths of hypo- and hyper-dense artifacts, attenuation within these, standard deviation in a region in proximity to the rod and in a reference region without presence of artifacts and the rod's diameter (*Table 2*) (3,12,18,19).

Artifact quantification algorithms: Part 1—automated methods

All algorithms were embedded in the Matlab-Environment (Mathworks, Natick, MA, USA). Images were processed in a fully automated fashion. First, images were converted from Hounsfield unit (HU) values to CT numbers by shifting from a range of -1,024 to 3,071 to a range of 0 to 4,095. The rod's center was identified by determining its

Table 1 Image acquisition and reconstruction parameters

Img#	Rod properties		Dose exposure					Pitch	Collimation (n × mm)	FOV (mm)	Reconstruction				
	Alloy	Diameter (mm)	kV _p	ms	mA	mAs	mGy				Algorithm	Denoising	VMI	Kernel	MAR
1	Titan Grade 2	5	80	1,706	17	29	1.2	0.435	16×0.625	350	Hybrid IR	6/6	No	B	No
2	Titan Grade 2	5	80	1,706	17	29	1.2	0.435	16×0.625	350	Model-based IR	3/3	No	ST	No
3	Titan Grade 4	10	80	1,727	44	76	3.1	0.435	16×0.625	350	Hybrid IR	0/6	No	YA	No
4	Titan Grade 4	10	80	1,727	44	76	3.1	0.435	16×0.625	350	Hybrid IR	0/6	No	YA	Yes
5	Titan Grade 4	10	80	1,727	44	76	3.1	0.435	16×0.625	350	Hybrid IR	6/6	No	C	No
6	Titan Grade 4	10	80	1,910	78	149	4.5	0.392	64×0.625	350	Hybrid IR	0/6	No	B	No
7	Titan Grade 4	10	80	1,910	78	149	4.5	0.392	64×0.625	350	Model-based IR	3/3	No	ST	No
8	Titan Grade 4	10	80	1,910	78	149	4.5	0.392	64×0.625	350	Model-based IR	3/3	No	ST	Yes
9	Titan Grade 4	10	80	1,911	157	300	9.0	0.392	64×0.625	350	Hybrid IR	0/6	No	B	No
10	Titan Grade 4	10	80	1,911	157	300	9.0	0.392	64×0.625	350	Model-based IR	3/3	No	ST	Yes
11	Titan Grade 4	10	80	1,911	157	300	9.0	0.392	64×0.625	350	Model-based IR	3/3	No	ST	No
12	Titan Grade 2	5	80	1,914	209	400	12.0	0.392	64×0.625	350	Hybrid IR	0/6	No	B	No
13	Titan Grade 2	5	80	1,914	209	400	12.0	0.392	64×0.625	350	Model-based IR	3/3	No	ST	No
14	Titan Grade 4	10	80	1,724	174	300	12.2	0.435	16×0.625	350	Hybrid IR	0/6	No	C	No
15	Titan Grade 4	10	80	1,724	174	300	12.2	0.435	16×0.625	350	Hybrid IR	6/6	No	B	No
16	Titan Grade 4	10	80	1,724	174	300	12.2	0.435	16×0.625	350	Hybrid IR	0/6	No	C	Yes
17	Titan Grade 4	10	120	1,910	78	149	13.5	0.392	64×0.625	350	Model-based IR	3/3	No	ST	Yes
18	Titan Grade 4	10	120	1,910	78	149	13.5	0.392	64×0.625	350	Spectral R	0/6	Yes	B	Yes
19	Titan Grade 2	5	120	1,835	109	200	21.2	0.409	32×0.625	350	Hybrid IR	1/6	No	B	No
20	Titan Grade 4	10	120	1,852	243	450	47.6	0.405	32×0.625	350	Hybrid IR	0/6	No	B	Yes
21	Titan Grade 4	10	120	1,852	243	450	47.6	0.405	32×0.625	350	Hybrid IR	0/6	No	B	No
22	Titan Grade 4	10	120	1,852	243	450	47.6	0.405	32×0.625	350	Spectral R	0/6	Yes	B	No
23	Titan Grade 2	5	140	1,914	105	201	26.0	0.392	64×0.625	350	Model-based IR	3/3	No	ST	Yes
24	Titan Grade 2	5	140	1,914	209	400	51.9	0.392	64×0.625	350	Model-based IR	1/3	No	ST	No
25	Titan Grade 2	5	140	1,914	209	400	51.9	0.392	64×0.625	350	Spectral R	3/6	Yes	B	No

For each image (img#), rod properties, dose-associated parameters, acquisition-associated parameters and reconstruction-associated parameters are reported. Kernel: B, ST—soft kernels; YA and C—sharp kernels. FOV, field of view; VMI, virtual monoenergetic images; MAR, metal artifact reduction postprocessing.

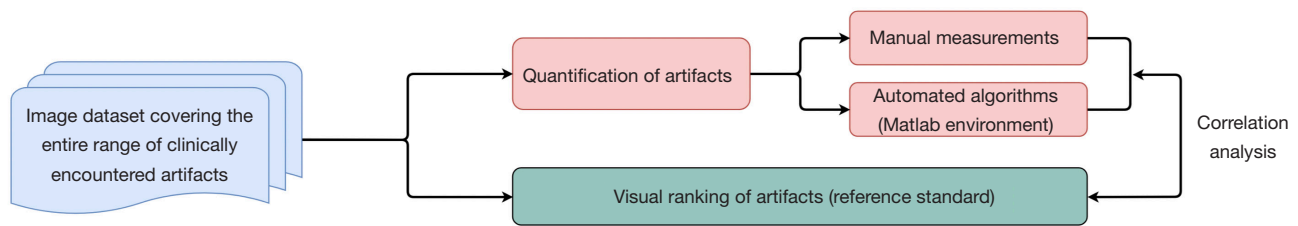


Figure 1 Study concept. The main idea of this study was to determine an artifact quantification method that correlates most closely to visual perception of artifacts.

center of gravity, further the rod's diameter was calculated. Several circular/annular regions were defined to which all algorithms were applied, radii were determined relative to the rods' radius (r_{rod} ; $1.5 \times r_{rod}$, $3.0 \times r_{rod}$ and $4.5 \times r_{rod}$). This resulted in a total of 3 circles defining a total of 3 annular areas. This approach was deliberately chosen to allow adaption to other objects and needs. Quantitative metrics were calculated along the circles as well as within the annular areas as described in the following sub-sections. Pixel values along the circle and within the annular areas were used; no interpolation was applied. *Table 2* provides an overview on used algorithms.

Standard deviation

The standard deviation along the circles and the annular area was calculated.

Fraction of hypodense and hyperdense artifacts

The average HU value within a reference region in absence of artifacts was determined and defined as the expected value (12). Within the annular regions, all pixels more than 2 and 3 standard deviations offset to these values were considered to have an artifact burden. Their fractional share was calculated relative to the absolute number of pixels within the area (%artifact 2σ and 3σ). In addition, artifacts were categorized either as hyperdense (%hyper, $\geq 2\sigma$ or $\geq 3\sigma$) or hypodense (%hypo, $\leq 2\sigma$ or $\leq 3\sigma$), again reported as a fractional share on all pixels.

Percent integrity uniformity (PIU)

The PIU method has been adapted from signal homogeneity considerations in electric signal processing. In imaging it has been used to describe image homogeneity in magnetic resonance (MR) imaging and has been proposed to allow for artifact quantification in CT imaging (5). In this context, it has been defined as $PIU = \left(1 - \frac{(S_{max} - S_{min})}{(S_{max} + S_{min})}\right) * 100\%$. A PIU of

100% indicates a perfectly homogeneous image without presence of artifacts. As proposed, PIU was calculated within the annular areas (5). As it initially has been proposed for (linear) electric signals, PIU was additionally calculated on the different circle circumferences. Two forms of PIU measurement were evaluated: one which used the minimum and maximum CT numbers as described and another which used 10th percentile and 90th percentile CT numbers, respectively.

Frequency domain methods

Previous studies have proposed artifact quantification within the frequency domain (16,17). This approach is based on observations that metal artifacts introduce oscillations, e.g., alternating “bright-dark” streaks, which are otherwise not present in an artifact-free image. In order to quantify these oscillations, a circle is selected which surrounds the artifact source. Pixels which lie along the circle are extracted in a counter-clockwise direction to produce a 1-dimensional signal of CT numbers. The signal is transformed into a spectrum of spatial frequencies content using the fast Fourier Transform. The magnitude of each Fourier coefficient was computed and binned by summing across the relevant frequencies. Multiple binning strategies were investigated: double bins (1–2, 3–4, and so on), binning in quarters (e.g., 1–10 for a spectrum with 40 Fourier coefficients), binning in half and one single bin.

Further details on image and signal processing are shown in *Figure 2*.

Evaluation of algorithms

The best suited method was considered to show a close correlation to the subjective ranking that served as gold standard in our study. To determine the closeness of fit Spearman's Tau was calculated, and only significant

Table 2 Methods for artifact quantification. Each method was applied to radii of 1.5-, 3.0- and 4.5-fold the rod's radius

Procedure	Analysis domain	Method	Subcategory	Analysis dimension	Analysis explanation	
Manual	Image	Rod radius	Radius (manual)	1-dimensional	Rod radius	
		Width	Hyperdense	1-dimensional	Width of hyperdense artifact/streak	
		Width	Hypodense	1-dimensional	Width of hypodense artifact/streak	
		HU	Hyperdense	2-dimensional (ROI-based)	HU within hyperdense artifact	
		HU	Hypodense	2-dimensional (ROI-based)	HU within hypodense artifact	
		SD	With artifacts	2-dimensional (ROI-based)	SD in tissue with visually perceived artifacts	
		SD	Without artifacts	2-dimensional (ROI-based)	SD in tissues without visual perceived artifacts	
		SDcorr	–	2-dimensional (ROI-based)	Corrected image noise (Δ SD w and w/o artifacts)	
Automated	Image	Rod radius	Radius (auto)	1-dimensional	Rod radius	
		PIU	Maximum pixel values	1-dimensional (circumference)	PIU based on min and max voxel value in linear circumference	
		PIU	10th/90th percentile	1-dimensional (circumference)	PIU based on 10th/90th percentile voxel value in linear circumference	
		SD	–	1-dimensional (circumference)	SD linear	
		% artifact	$2\sigma/3\sigma$	2-dimensional (annular region)	Share of all artifact pixels based on $2\sigma/3\sigma$ -threshold	
		% hyperdense A	$2\sigma/3\sigma$	2-dimensional (annular region)	Share of hyperdense pixels based on $2\sigma/3\sigma$ -threshold	
		% hypodense A	$2\sigma/3\sigma$	2-dimensional (annular region)	Share of hypodense pixels based on $2\sigma/3\sigma$ -threshold	
		PIU	Maximum pixel values	2-dimensional (annular region)	PIU based on min and max voxel value in annular area	
		PIU	10th/90th percentile	2-dimensional (annular region)	PIU based on 10th/90th percentile voxel value in annular area	
		SD	–	2-dimensional (annular region)	SD area	
	Frequency	FFT		Single 1–2, 3–4, ...	1-dimensional (circumference)	FFT small frequency bins 1–2, 3–4, ...
				No binning	1-dimensional (circumference)	FFT no binning
				Binning in lower and upper half	1-dimensional (circumference)	FFT upper/lower half frequency bins
				Binning in quarters	1-dimensional (circumference)	FFT half frequency binned by quarters

Comprehensive list of all evaluated artifact quantification algorithms. Further details, including mention of groups first describing the different methods are listed in the manuscript body. HU, Hounsfield unit; SD, standard deviation; PIU, percent integrity uniformity; FFT, fast Fourier Transform; ROI, region of interest.

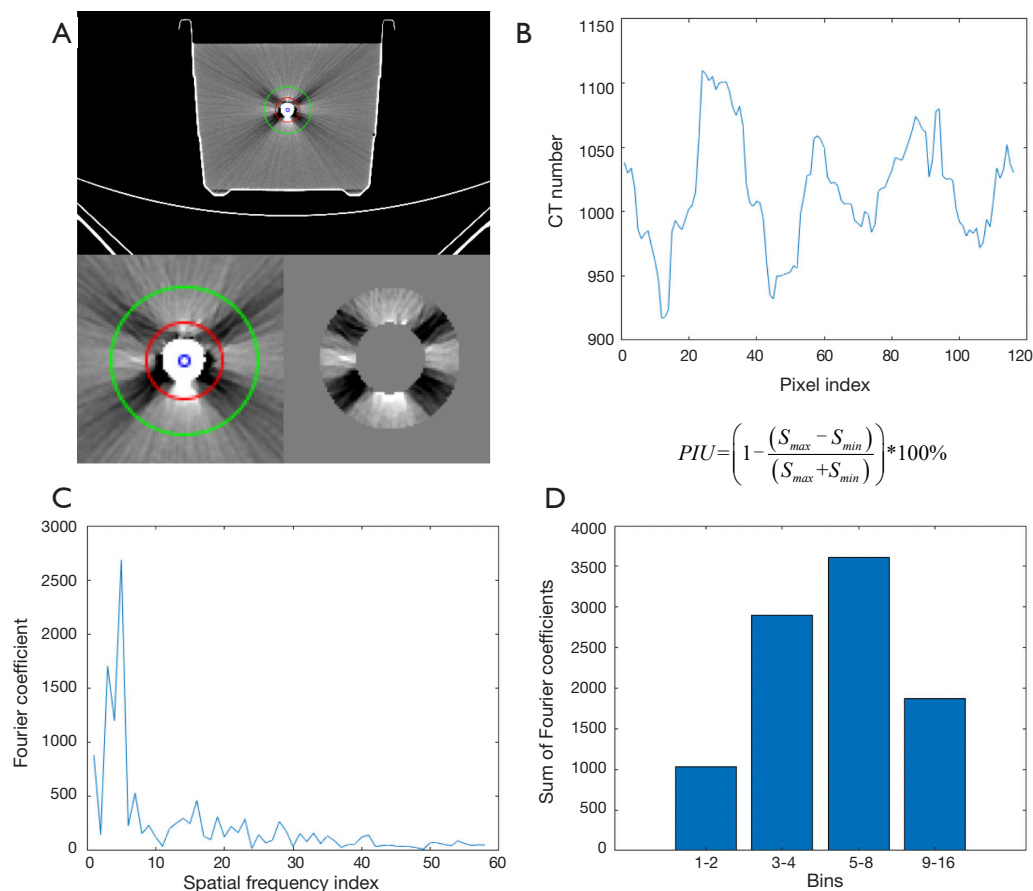


Figure 2 Automated quantification methods. Example phantom with rod insert giving significant metal artifact (A). Blue circle denotes the rod center, red circle denotes a radius of $1.5 \times$ rod's radius ($1.5\times$) from the rod center, green circle denotes a radius of $3 \times$ rod's radius from the rod center. (B) CT number values along the $1.5\times$ circle are plotted starting from 3 o'clock and proceeding in a counter-clockwise direction. Peaks and valleys in the curve can be seen corresponding to the hyper- and hypo-intense streaks seen in (A). The minimum and maximum values (or the 10th/90th percentile values) are used for quantification based on the PIU using the formula displayed. PIU from this curve is 90.48%. (C) The fast Fourier Transform (FFT) of the curve in (B) is taken to obtain the spatial frequency content of the signal. The magnitude of the Fourier coefficients is plotted starting from the zero-frequency. This curve is used for the FFT-based metal artifact quantification. Since the curve contains predominantly low-frequency information in this data set, most of the Fourier coefficients are contained in the lower frequencies. The Fourier coefficients can be binned in different ways. (D) Here, example bins are illustrated (D). Values are summed to produce the four bin values shown, corresponding to the first and second Fourier coefficients, third and fourth coefficients, fifth through eighth coefficients, and ninth through sixteenth coefficients, respectively. These bins were applied in a previous report (16).

correlations were considered in the further analysis. All statistic procedures were carried out using JMP software (SAS Institute, Cary, NC, USA).

Results

Regarding the reference ranking, intra- and inter-reader agreement as depicted by ICC were 0.856–0.928 and 0.875, respectively. Reference order of images is displayed in

Figure 3.

Regarding the manual measurements, inter-reader reliability ranged from 0.247 and 0.986 for the different items. Intra-reader reliability was higher, ranging from 0.497 to 0.993. Best reliability for both, intra- and inter-reader, was observed for 1-dimensional measurements of the rod's diameter. Poorest reliability was found in measurements of the artifact width (0.247–0.862, Table 3). The automated algorithms were applied to the images twice showing

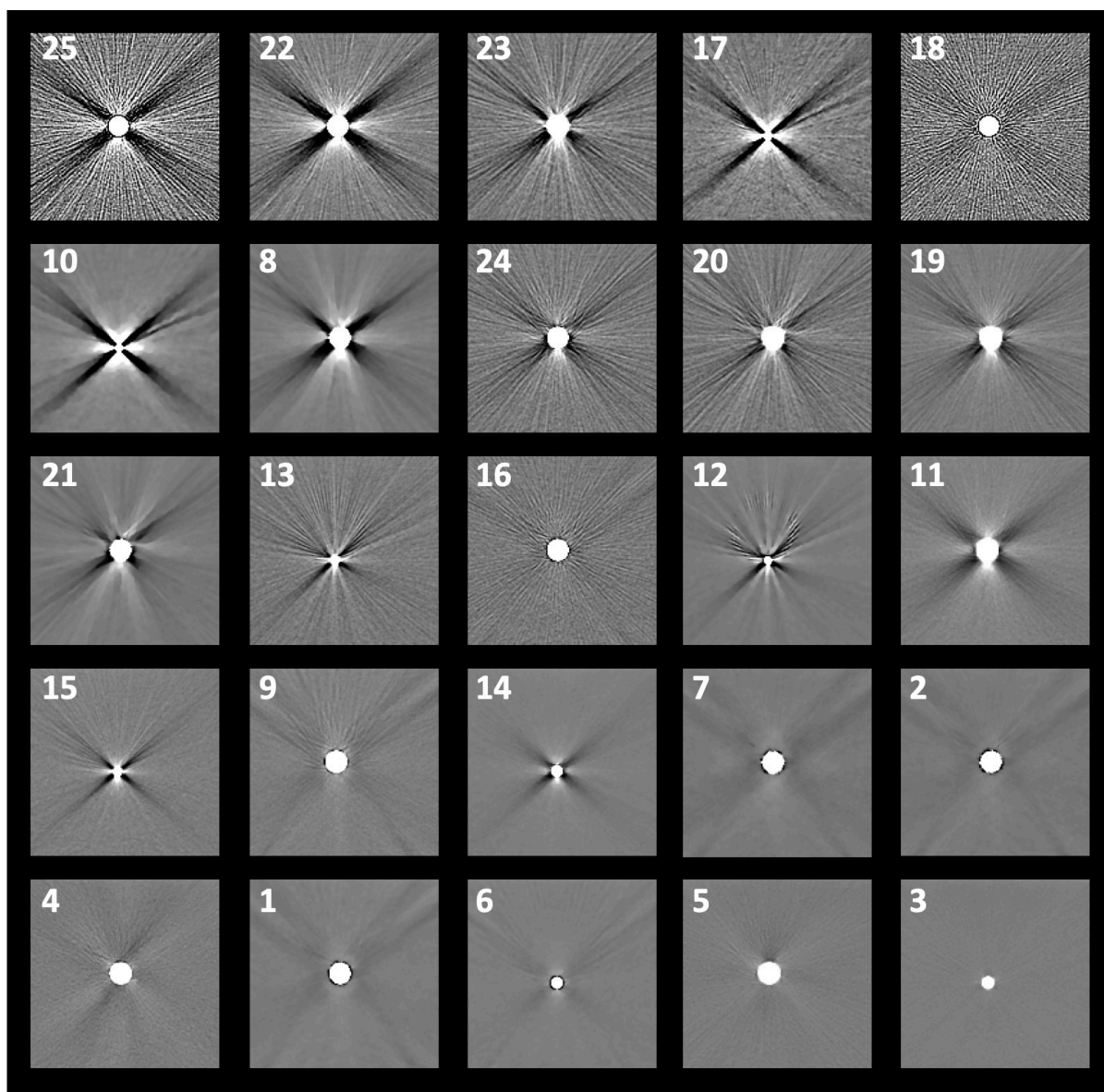


Figure 3 Reference ranking. Reference ranking obtained from visual analysis. Images are ordered from most artifact to least artifact impaired image (top left to bottom right). Further, the Img# is displayed, see *Table 1*.

perfect reproducibility.

No method was capable to order all images in the reference order. Best correlation as indicated by Spearman's Tau was 0.977 obtained for a quantification algorithm within the frequency domain in a 1-dimensional circular line around the rod only considering voxels in the lower frequency bin (lowest quarter of all bins). Yet, even this (best) algorithm ranked 15/25 wrong; however, considering only differences greater than one rank, only 5 images were

ranked incorrectly (*Figure 4*). For the frequency domain-based methods, various binning methods were included in the final analysis. Here, bins in the lower end showed best performance.

Automated measurement of standard deviation within a circular ROI demonstrated a correlation of 0.729, yet this resulted in misranking of 24/25 images (of which 22/25 were misranked for more than one rank).

Automated analysis within the image domain in terms

Table 3 Reliability measures. Intraclass correlation coefficient for the different items is reported

Item	Inter-reader agreement	Intra-reader agreement (averaged for all readers)
Reference Reading	0.875	0.886
Manual measurements		
HU hyperdense	0.844	0.843
HU hypodense	0.751	0.818
Rod diameter	0.986	0.991
SD with artifacts	0.678	0.817
SD without artifacts	0.934	0.929
Width hyperdense	0.247	0.560
Width hypodense	0.360	0.729

All manual measurements are based on regions of interest. HU, Hounsfield unit (measurement of artifact attenuation); SD, measurement of standard deviation based on regions of interest.

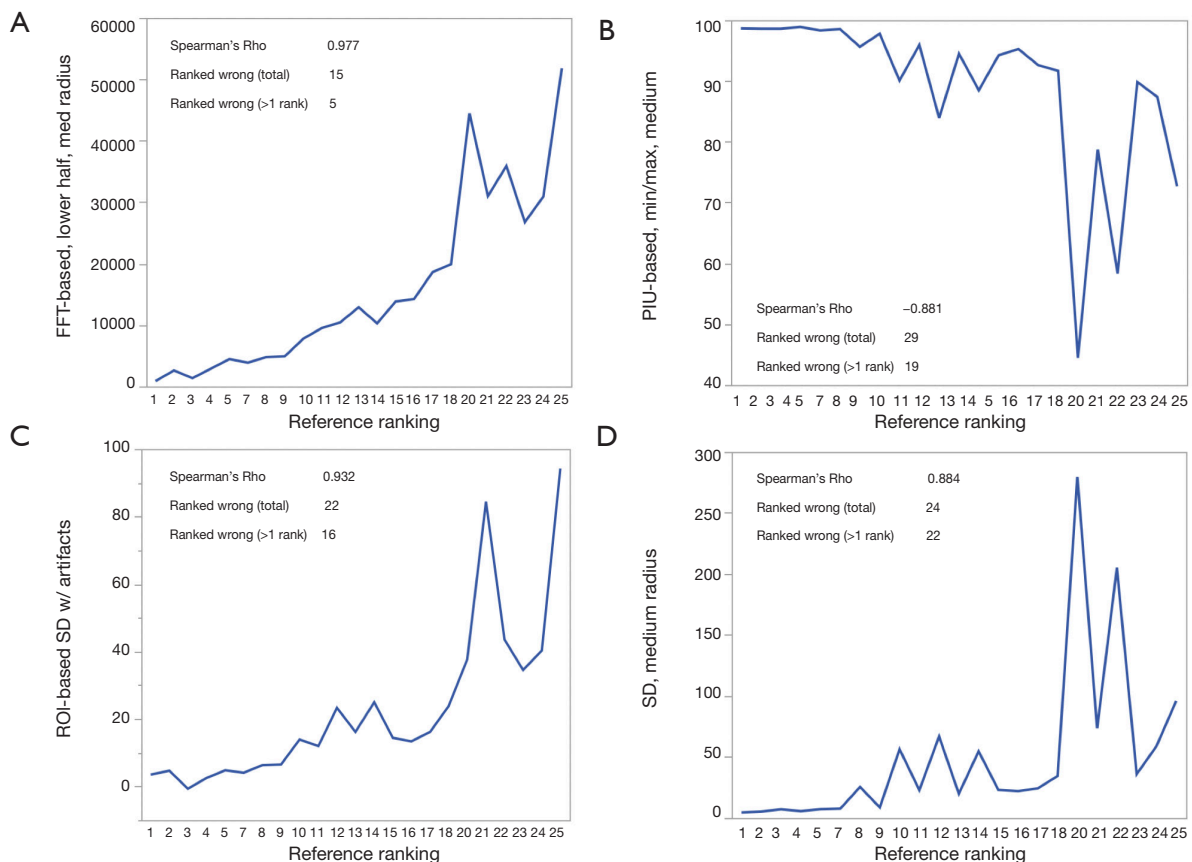


Figure 4 Best methods for artifact quantification. Demonstrating of the best performing quantification algorithm from each category. The FFT-based analysis after binning lower half images with a medium radius demonstrate closest agreement with visual artifact perception (Spearman's Rho =0.977) (A). Further, best method based on percent integrity uniformity analysis (PIU) (B), manual measurements based on regions of interest (ROI) (C) and automated determination of standard deviation (SD) (D), are shown. FFT, fast Fourier Transform.

of PIU yielded a correlation of -0.881 corresponding to 23 and 19 misclassified images (overall and more than one rank, respectively). Interestingly, the area-based methods within the image domain yielded a correlation of 0.576 only. Here, 22/25 images were ranked incorrect, of these 20 received a rank more than one apart from their actual one.

Regarding different sizes of circular or annual ROI around the rod, the larger one's demonstrated to correlate better with visual artifact perception, irrespective of quantification algorithm. Correlation coefficients for all methods and radii are available in an electronic supplement (ESM 1, table online: <http://fp.amegroups.cn/cms/b58190d41af19ad6958fabc23822874c/qims.2020.04.03-1.pdf>).

The best manual measurement (measurement of standard deviation in areas with visually perceived artifacts) misranked 22/25 images of which 16 were misclassified more than 2 ranks.

Discussion

The aim of this study was to evaluate different methods for quantitative assessment of artifacts occurring due to metal in CT imaging. Several approaches were identified in literature and compared to visual artifact perception. Yet, none of these were capable to correctly rank artifacts based on quantitative parameters in the order they are visually perceived. In addition, manual measurements demonstrated limited reproducibility (intra- and inter-reader). The identified methods that showed closest agreement with visual artifact perception included image analysis within the frequency domain.

Artifacts due to metal hardware are a fairly common problem in CT imaging. Hence, several methods have been proposed to reduce these. During the past years, dedicated algorithms have been introduced by all major vendors. These algorithms aim to specifically identify artifact-impaired voxels within the projection or reconstruction-domain and to reduce these using an iterative loop (5,20,21). Another more recent method that is available to reduce artifacts in CT imaging are virtual monoenergetic images from dual energy CT (14,19,22,23). As suggested by recent developments, numerous studies in the field of metal artifact reduction in CT imaging are available. As of April 2019, approximately 1,300 studies are available that contain the search term combination “computed tomography” and “artifact” (24). Yet, the comparability between different studies and groups is challenging as different measures are

considered for artifact quantification.

Most of these studies emphasize on a visual assessment of artifact extent as standard of reference. This appears valid, as the aim of most metal artifact reduction techniques is to improve diagnostic assessment in the reading room after all. In addition, the majority of studies include a quantitative portion. For the latter, suggestions vary from simple manual measurements to advanced methods for quantitative analysis. At first instance, methods can be subdivided into automated and manual methods. For the latter, several challenges need to be acknowledged. First, one common problem in manual image analysis is reader bias. Particularly in studies comparing two different means for image reconstruction with respect to their artifact extent, a blinding of readers is often not sufficiently possible. This is due to the fact, that the visual appearance of images is very different when it comes to iteratively reconstructed images or virtual monoenergetic images. We demonstrated, that (even in specifically instructed readers), intra- and interreader reliability or agreement is limited. Hence, it is very difficult to compare studies between different groups based on manual measurements.

The manual method showing best correspondence with visual artifact perception is based on the measurement of standard deviation. It appears rational, to conduct such measurements as either streak (hyper- or hypo-dense) would result in an increase of image noise; however, nonlinear means of noise reduction as inherent to iterative image reconstruction techniques may extensively alter noise characteristics. Große Hokamp *et al.* suggested to use the so-called corrected image noise instead (19,25,26). This parameter is obtained by subtraction of standard deviation in areas without presence of artifacts from standard deviation in areas with presence of artifact and is thought to only quantify noise due to artifacts (25). Yet, the problem of limited reproducibility remains.

The herewith proposed automated methods, on the other hand, hold the inherent advantage of exact reproducibility. The method that showed best performance carries out the computation within the frequency domain and has been proposed by Mangold *et al.* (16); while the initial study suggested binning frequencies in an arbitrary fashion, we evaluated different systematic binning approaches. In line with the aforementioned study, lower frequency bins demonstrated best correspondence with visual artifact perception.

As stated above, visual perception is considered gold

standard for evaluation of image artifacts. To evaluate the different methods, we were in need of some sort of reference database which we obtained from a two-pair forced choice analysis. As opposed to other established scales, such as Likert-scales, this method forces the reader to decide on the less artifact impaired image (27). Although other factors (reconstruction algorithm, kernel) may lead to an additional bias; the excellent intra- and inter-reader agreement suggest that this approach is legitimate for our purpose.

Other imaging parameters, e.g., the signal- or contrast-to-noise ratio, are commonly accepted and reported in scientific studies. These parameters therefore allow for comparison of results between different studies. Such comparability as well as the associated reproducibility is inherently linked to the idea of good scientific practice. Therefore, the code for implementation within the Matlab-Environment is made available online (ESM 2, available online: <http://fp.amegroups.cn/cms/f64c5f1bfc1bbb0fe7c06bbc1185b5a7/qims.2020.04.03-2.txt>) and encourage its utilization in future studies addressing metal artifact reduction in CT studies.

There are some other limitations that need to be addressed or considered. First, we only included a fairly limited number of images. This became necessary as the number of comparisons using 2-pair forced choice method grows quadratically. Second, all scans were conducted in a homogeneous phantom and their applicability in an anatomic setting might be more limited. Further, all images were acquired using a single scanner; however, image acquisition and reconstruction parameters covered a very wide spectrum of possible combinations. We used a customized phantom containing only one type of medically used metal alloy (Titanium). While a commercially available setup would allow for comparability, we feel that this dedicated phantom represents reality closely and also enabled us to include differently sized rods. Last, none of the methods worked perfectly; therefore, quantitative analysis should always be supplemented by visual analysis.

Quantifying artifacts due to metal hardware is challenging. We propose one method that allows for automated quantification of such artifacts using a Fourier transformed linear ROI a lower end frequency bins. The code for implementation within the Matlab-Environment is made available as an electronic supplement. We encourage utilizing this method over other earlier suggested approaches and particularly over manual measurements as it shows a close agreement to visual artifact perception and is not limited by inter- or intra-reader agreement.

Acknowledgments

Funding: Parts of this study have been funded by the Koeln Fortune Program/Faculty of Medicine, University of Cologne (339/2018 to N Große Hokamp).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/qims.2020.04.03>). NGH reports grants and personal fees from Philips Healthcare, outside the submitted work. BE reports personal fees from BioInvision, Inc., outside the submitted work. DPS reports personal fees from Cook Medical, personal fees from mediaire GmbH, outside the submitted work. JAH reports grants from Philips, outside the submitted work. DM reports personal fees from Philips Healthcare, outside the submitted work. The other authors have no conflicts of interest to declare.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Wu M, Keil A, Constantin D, Star-Lack J, Zhu L, Fahrig R. Metal artifact correction for x-ray computed tomography using kV and selective MV imaging. *Med Phys* 2014;41:121910.
2. Boas FE, Fleischmann D. CT artifacts: causes and reduction techniques. *Imaging Med* 2012;4:229-40.
3. Neuhaus V, Große Hokamp N, Abdullayev N, Rau R, Mpotsaris A, Maintz D, Borggrefe J. Metal artifact reduction by dual-layer computed tomography using virtual monoenergetic images. *Eur J Radiol* 2017;93:143-8.
4. Buckwalter KA, Parr JA, Choplin RH, Capello WN. Multichannel CT Imaging of Orthopedic Hardware and Implants. *Semin Musculoskelet Radiol* 2006;10:86-97.
5. Große Hokamp N, Hellerbach A, Gierich A, Jordan DW, Visser-Vandewalle V, Maintz D, Haneder S. Reduction of

- Artifacts Caused by Deep Brain Stimulating Electrodes in Cranial Computed Tomography Imaging by Means of Virtual Monoenergetic Images, Metal Artifact Reduction Algorithms, and Their Combination. *Invest Radiol* 2018;53:424-31.
6. Kuchenbecker S, Faby S, Sawall S, Lell M, Kachelrieß M. Dual energy CT: How well can pseudo-monochromatic imaging reduce metal artifacts? *Med Phys* 2015;42:1023-36.
 7. Lee YH, Park KK, Song HT, Kim S, Suh JS. Metal artefact reduction in gemstone spectral imaging dual-energy CT with and without metal artefact reduction software. *Eur Radiol* 2012;22:1331-40.
 8. Dong Y, Shi AJ, Wu JL, Wang RX, Sun LF, Liu AL, Liu YJ. Metal artifact reduction using virtual monochromatic images for patients with pedicle screws implants on CT. *Eur Spine J* 2016;25:1754-63.
 9. Wellenberg RH, Boomsma MF, van Osch JA, Vlassenbroek A, Milles J, Edens MA, Streekstra GJ, Slump CH, Maas M. Quantifying metal artefact reduction using virtual monochromatic dual-layer detector spectral CT imaging in unilateral and bilateral total hip prostheses. *Eur J Radiol* 2017;88:61-70.
 10. Kidoh M, Nakaura T, Nakamura S, Tokuyasu S, Osakabe H, Harada K, Yamashita Y. Reduction of dental metallic artefacts in CT: value of a newly developed algorithm for metal artefact reduction (O-MAR). *Clin Radiol* 2014;69:e11-6.
 11. Huang JY, Kerns JR, Nute JL, Liu X, Balter PA, Stingo FC, Followill DS, Mirkovic D, Howell RM, Kry SF. An evaluation of three commercially available metal artifact reduction methods for CT imaging. *Phys Med Biol* 2015;60:1047-67.
 12. Boomsma MF, Warringa N, Edens MA, Mueller D, Ettema HB, Verheyen CC, Maas M. Quantitative analysis of orthopedic metal artefact reduction in 64-slice computed tomography scans in large head metal-on-metal total hip replacement, a phantom study. *Springerplus* 2016;5:405.
 13. Guggenberger R, Winklhofer S, Osterhoff G, Wanner GA, Fortunati M, Andreisek G, Alkadhi H, Stolzmann P. Metallic artefact reduction with monoenergetic dual-energy CT: systematic ex vivo evaluation of posterior spinal fusion implants from various vendors and different spine levels. *Eur Radiol* 2012;22:2357-64.
 14. Bamberg F, Dierks A, Nikolaou K, Reiser MF, Becker CR, Johnson TRC. Metal artifact reduction by dual energy computed tomography using monoenergetic extrapolation. *Eur Radiol* 2011;21:1424-9.
 15. Haneder S, Siedek F, Doerner J, Pahn G, Grosse Hokamp N, Maintz D, Wybranski C. Thoracic-abdominal imaging with a novel dual-layer spectral detector CT: intra-individual comparison of image quality and radiation dose with 128-row single-energy acquisition. *Acta Radiol* 2018;59:1458-65.
 16. Mangold S, Gatidis S, Luz O, König B, Schabel C, Bongers MN, Flohr TG, Claussen CD, Thomas C. Single-source dual-energy computed tomography: use of monoenergetic extrapolation for a reduction of metal artifacts. *Invest Radiol* 2014;49:788-93.
 17. Aissa J, Boos J, Schleich C, Sedlmair M, Krzymyk K, Kröpil P, Antoch G, Thomas C. Metal Artifact Reduction in Computed Tomography After Deep Brain Stimulation Electrode Placement Using Iterative Reconstructions. *Invest Radiol* 2017;52:18-22.
 18. Laukamp KR, Lennartz S, Neuhaus VF, Große Hokamp N, Rau R, Le Blanc M, Abdullayev N, Mpotsaris A, Maintz D, Borggrefe J. CT metal artifacts in patients with total hip replacements: for artifact reduction monoenergetic reconstructions and post-processing algorithms are both efficient but not similar. *Eur Radiol* 2018;28:4524-33.
 19. Große Hokamp N, Neuhaus V, Abdullayev N, Laukamp K, Lennartz S, Mpotsaris A, Borggrefe J. Reduction of artifacts caused by orthopedic hardware in the spine in spectral detector CT examinations using virtual monoenergetic image reconstructions and metal-artifact-reduction algorithms. *Skeletal Radiol* 2018;47:195-201.
 20. Willeminck MJ, Leiner T, de Jong PA, de Heer LM, Nievelstein RA, Schilham AM, Budde RP. Iterative reconstruction techniques for computed tomography part 2: initial results in dose reduction and image quality. *Eur Radiol* 2013;23:1632-42.
 21. Willeminck MJ, Noël PB. The evolution of image reconstruction for CT—from filtered back projection to artificial intelligence. *Eur Radiol* 2019;29:2185-95.
 22. D'Angelo T, Cicero G, Mazziotti S, Ascenti G, Albrecht MH, Martin SS, Othman AE, Vogl TJ, Wichmann JL. Dual energy computed tomography virtual monoenergetic imaging: technique and clinical applications. *Br J Radiol* 2019;92:20180546.
 23. Yu L, Leng S, McCollough CH. Dual-energy CT-based monochromatic imaging. *AJR Am J Roentgenol* 2012;199:S9-15.
 24. US National Library of Medicine and National Institutes of Health. Available online: <https://pubmed.ncbi.nlm.nih.gov>
 25. Große Hokamp N, Laukamp KR, Lennartz S, Zopf

- D, Abdullayev N, Neuhaus VF, Maintz D, Borggreffe J. Artifact reduction from dental implants using virtual monoenergetic reconstructions from novel spectral detector CT. *Eur J Radiol* 2018;104:136-42.
26. Laukamp KR, Zopfs D, Lennartz S, Pennig L, Maintz D, Borggreffe J, Große Hokamp N. Metal artifacts in patients with large dental implants and bridges: combination of metal artifact reduction algorithms and virtual monoenergetic images provides an approach to handle even strongest artifacts. *Eur Radiol* 2019;29:4228-38.
27. Ellmann S, Kammerer F, Brand M, Allmendinger T, May MS, Uder M, Lell MM, Kramer M. A Novel Pairwise Comparison-Based Method to Determine Radiation Dose Reduction Potentials of Iterative Reconstruction Algorithms, Exemplified Through Circle of Willis Computed Tomography Angiography. *Invest Radiol* 2016;51:331-9.

Cite this article as: Große Hokamp N, Eck B, Siedek F, Pinto dos Santos D, Holz JA, Maintz D, Haneder S. Quantification of metal artifacts in computed tomography: methodological considerations. *Quant Imaging Med Surg* 2020;10(5):1033-1044. doi: 10.21037/qims.2020.04.03