This section explains the additional technical details on the methods that we used for quantitative and qualitative evaluation of reduced and standard dose CT scans. Five quantitative metrics were used to analyze the influence of radiation exposure in the perception of CT images at organ and tissue levels. These metrics perform objective comparisons based on both organ segmentation and image intensity features. While SSIM and GMSD provide quality assessment based on similarities of luminance, contrast, texture, and edge information, DSC, HD, and WEST evaluate the similarity between two region/volume delineations. Figure A1 illustrates the conventional CT attenuation ranges for different tissue types, which we used them for tissue level comparisons in our experiments by using probability density functions derived from histograms of those tissue densities after tissue-specific thresholding of CT images.

**Dice Similarity Coefficient (DSC):** DSC [30] evaluates similarity between two segmented objects based on their overlap regions as shown in Figure A2-A, given two segmented regions, $X$ and $Y$, it is defined as

$$DSC(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|}.$$
Here, $|X \cap Y|$ represents the volume of the overlapped region between $X$ and $Y$ (green region), which is normalized by the total volume of $X$ and $Y$. Two identical objects with perfect overlap will have a DSC of 1, and two objects with no overlap will have a DSC of 0.

In our experiments, two experts performed segmentation of lung, liver, and spleen, blinded to each other’s delineations. The DSC was calculated between registered organ segmentation for standard dose and reduced dose scans. The organs under investigation do not have disease diagnosed; hence, deviation from DSC from one scan to another indicates differences in the image quality. Based on this, the organ segmentation would have high DSC if dose posed minor influence over the experts’ perception of organ boundaries. One limitation of DSC is that it only quantifies the overlapping property without considering the boundary matching.

For instance, as shown in Figure A2.B and C, the two cases have similar DSC; however, we usually prefer multiple small discrepancies (B) to a major miss (C). Therefore, we need other measurements to avoid such misleading and be complementary to DSC.

**Hausdorff Distance Coefficient (HD):** HD [31] measures boundary mismatches between two objects. It is considered as a complementary metric to DSC. Let $x \in X, y \in Y$ denote the boundary voxels of objects $X$ and $Y$, then HD is defined as

$$HD(X, Y) = \{\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y)\},$$
where \(d(x, y)\) measures the Euclidean distance between two boundary voxel locations. For a given boundary location, \(\inf_{y \in Y} d(x, y)\) finds the distance between \(x\) and its closest correspondence in \(Y\), and \(\sup_{x \in X} \inf_{y \in Y} d(x, y)\) determines the maximum distances among all boundary points in \(X\). Similarly, \(\sup_{y \in Y} \inf_{x \in X} d(x, y)\) finds the maximum distances among all boundary points in \(Y\) to their closest correspondences in \(X\) (as illustrated in Figure A2.A). Therefore, \(HD(X, Y)\) measures the maximum distance between two object boundaries, giving a summary of how well two boundary shapes overlapped. Two identical objects with perfect alignment will have a HD value of 0, and the more mismatches between two boundaries, the greater HD value will be. In our experiment, HD is calculated between segmented organs obtained from both standard and reduced dose CT scans. Note that HD complements DSC by considering the distance between two boundaries, but it has its own limitations too. As in Figure A2.D and E, the two cases are almost identical, but a single outlier included in (E) leads to quite significant differences in HD computation. Hence, instead of only considering the greatest distance, more sophisticated measurements are needed to evaluate shape dissimilarity as outlined below.

**Weighted Spectral Distance (WESD):** WESD [34] serves as an alternative measure to HD for identifying shape dissimilarity between objects. It compares the similarity of two objects, \(X\) and \(Y\), according to overall geometrical information. Given the piecewise smooth object boundaries \(\Omega_X, \Omega_Y\) of \(X\) and \(Y\), WESD and normalized WESD (nWESD) are defined as
\[ W_{E(S)D}(X, Y) = \left[ \sum_{n=1}^{N} \left( \frac{|X_n - Y_n|}{X_n Y_n} \right)^p \right]^{1/p} \]

and

\[ nW_{E(S)D}(X, Y) = \frac{W_{E(S)D}(X, Y)}{W(X, Y)} \]

where \( W \) is the shape dependent normalization factor, \( X_n \) and \( Y_n \) are the eigenvalue sequence (spectrum) of \( \Omega_X \) and \( \Omega_Y \), \( N \) is the number of eigenvalue truncated, and \( p \) controls the sensitivity with respect to shape differences at finer scales. \( W_{E(S)D} \) is proven to be theoretically sound and have practical benefits for medical image analysis applications. It provides finer control over the evaluation of shape similarity rather than simple boundary distance given by HD. It is mapped to the \([0, 1)\) interval with 0 representing identical objects. For our experiments, it is calculated in the same manner as DSC and HD with two object delineations obtained from standard and reduced dose CT scans.

**Structural Similarity Index (SSIM):** SSIM [33] measures the image quality difference between two images based on similarities of local luminance, contrast, and structure/texture information. Spatially close pixels have strong inter-dependencies; hence, they carry important information about the structure of the objects in the visual scene. Given two images \( I \) and \( J \), and two corresponding region definition \( R_x \) and \( R_y \), SSIM is defined as
\[ SSIM(I, J) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \]

where \(\mu_x, \mu_y, \sigma_x, \sigma_y\) and \(\sigma_{xy}\) are the average, variance and covariance of the intensities within windows pertaining to the images I and J, respectively. SSIM values between -1 to 1, with 1 representing two identical images. In our experiments, we derive organ based SSIM by computing SSIM between segmented organs from standard and reduced dose CT scans. The resulting SSIM values demonstrate the image quality in structure perception for the organs segmented from standard and reduced dose CT scans.

**Gradient Magnitude Similarity Deviation (GMSD):** GMSD [32] uses the global variation of gradient-based quality map for overall image quality prediction. Compared with SSIM, it provides more global information by comparing the edge details contained within two images. GMSD is a variation of the Gradient Magnitude Similarity (GMS) map defined at a voxel location \(i\) between two images I and J (with in total K voxels) as

\[ GMS_i(I, J) = \frac{(2m_i(i)m_j(i) + c)}{(m_i^2(i)+m_j^2(i) + c)} \]

where \(m_i(i)\) and \(m_j(i)\) are the gradient magnitudes of the images, c is a positive constant for numerical stability. GMSD can then be represented as
\[ GMSD(I, J) = \sqrt{\frac{1}{K} \sum_{i=1}^{K} \left( GMS_i(I, J) - \frac{1}{K} \sum_{j=1}^{K} GMS_j(I, J) \right)^2}. \]

GMSD ranges from 0 to 1 and, it reaches to 1 for two identical images. In our experiments, GMSD is calculated between each segmented organ, and resulting GMSD values were used to quantify the image quality change in gradient magnitude between standard and reduced dose CT scans.

**Results for individual measurements:** Figure A3 demonstrates the results of the above measurements in our experiments. Ten cases were randomly chosen as illustrative examples. As shown, the results were divided to 3x5 blocks whose rows represent five measurements of DSC, HD, WESD, SSIM, and GMSD; and columns represent three organs of heart, liver, and spleen. Each block consists of 4x10 numbers whose columns represent 10 cases, and rows represent four kinds of comparison pair configurations. As illustrated, for different similarity measurements, the global shape similarity measured by DSC, HD, and WESD between standard and reduced dose CT images are comparable to that of inter-observer agreement for the same image (first three row of blocks); while a difference can be observed for local appearance measurements of SSIM and GMSD.
Figures

Figure A1. CT attenuation ranges for different tissue types.

Figure A2. (A) HD metric is described as the largest shortest distances between two segmented...
boundaries. DSC captured the overlapping area (green) normalized by the total of two areas. (B, C, D, E) Illustrations of how the three complement each other: B has small but repetitive misses, C has a major miss with small general underestimation in other parts, D has slightly larger general underestimation, and E has similar general underestimation as D with a minor but significant miss. The DSC won't be able to differentiate B, C, D, E since all will have similar values. HD is capable of capturing the difference between (B, D - small) vs. (C, E – large). Furthermore, WESD can differentiate B vs. D, and C vs. E based on the shape of the two corresponding boundaries.

Figure A3. Results for five individual measurements used in our experiments. Five measurements have been computed for three organs (heart, liver, and spleen) between four pairs of segmented regions: two observers’ results for standard dose CT image (interSD), two observers’ results for reduced dose CT image (interLD), observer 1’s and observer 2’ results
for standard and reduced dose images (LDSD1 and LDSD2), respectively. The numbers were
color-coded and shown with regard to their corresponding color-bar.

REFERENCES

30. Dice, L. R., The selection index and its test of significance. Evolution, 1949. 3(3): p. 262-5.
31. Taha, A. A. and A. Hanbury, An Efficient Algorithm for Calculating the Exact Hausdorff
    Distance. IEEE Trans Pattern Anal Mach Intell, 2015. 37(11): p. 2153-63.
32. Xue, W., et al., Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual
    Image Quality Index. IEEE Trans Image Process, 2014. 23(2): p. 684-95.
33. Wang, Z., et al., Image quality assessment: from error visibility to structural similarity.
    IEEE Trans Image Process, 2004. 13(4): p. 600-12.
34. Konukoglu, E., et al., WESD-Weighted Spectral Distance for Measuring Shape
    Dissimilarity. Ieee Transactions on Pattern Analysis and Machine Intelligence, 2013.
    35(9): p. 2284-2297.