Published in final edited form as:

IEEE Trans Biomed Eng. 2015 December ; 62(12): 2812–2827. doi:10.1109/TBME.2015.2445054.

Seamless Insertion of Pulmonary Nodules in Chest CT Images

Aria Pezeshk,

Division of Imaging, Diagnostics, and Software Reliability (DIDSR), Office of Science and Engineering Laboratories (OSEL), Center for Devices and Radiological Health (CDRH), U.S. Food and Drug Administration (FDA), Silver Spring, MD 20993 USA

Berkman Sahiner,

DIDSR/OSEL/CDRH/FDA, Silver Spring, MD 20993 USA

Rongping Zeng,

DIDSR/OSEL/CDRH/FDA, Silver Spring, MD 20993 USA

Adam Wunderlich,

DIDSR/OSEL/CDRH/FDA, Silver Spring, MD 20993 USA

Weijie Chen, and

DIDSR/OSEL/CDRH/FDA, Silver Spring, MD 20993 USA

Nicholas Petrick

DIDSR/OSEL/CDRH/FDA, Silver Spring, MD 20993 USA

Abstract

The availability of large medical image datasets is critical in many applications such as training and testing of computer aided diagnosis (CAD) systems, evaluation of segmentation algorithms, and conducting perceptual studies. However, collection of data and establishment of ground truth for medical images are both costly and difficult. To address this problem, we are developing an image blending tool that allows users to modify or supplement existing datasets by seamlessly inserting a lesion extracted from a source image into a target image. In this study we focus on the application of this tool to pulmonary nodules in chest CT exams. We minimize the impact of user skill on the perceived quality of the composite image by limiting user involvement to two simple steps: the user first draws a casual boundary around a nodule in the source, and then selects the center of desired insertion area in the target. We demonstrate the performance of our system on clinical samples, and report the results of a reader study evaluating the realism of inserted nodules compared to clinical nodules. We further evaluate our image blending techniques using phantoms simulated under different noise levels and reconstruction filters. Specifically, we compute the area under the ROC curve (AUC) of the Hotelling observer (HO) and noise power spectrum (NPS) of regions of interest enclosing native and inserted nodules, and compare the detectability, noise texture, and noise magnitude of inserted and native nodules. Our results indicate the viability of our approach for insertion of pulmonary nodules in clinical CT images.

Index Terms

Poisson editing; pulmonary nodules; data augmentation; image composition

I. Introduction

MANY studies involving medical images require a large number of samples that have been validated or characterized by clinical experts. Some examples include evaluation of search tasks in perceptual studies or segmentation and volumetry algorithms where the true location and precise boundaries of lesions need to be known *a priori* and in large numbers in order to arrive at statistically significant conclusions. Computer aided diagnosis (CAD) systems comprise another important example. These systems have shown a high potential in assisting radiologists with the interpretation of the large amount of data produced by state-of-the-art medical imaging devices [1], [2], [3]. However, both the development and validation of CAD systems are hindered by the high cost and difficulties associated with the collection of large image repositories and the establishment of ground truth for these datasets. In terms of development, it is well known that the performance of any classifier is heavily affected by the size and representativeness of the data used for its training and CAD systems are no exception to this general rule [4], [5]. As for validation, both the standalone assessment of an individual CAD system and the comparison of a group of systems require a large number of representative samples [6], [7].

The availability of small and/or unbalanced training sets is a common problem in many classification tasks. One of the most established approaches to overcoming this problem has been to boost the size of a training set by adding synthetically generated samples. The mechanisms used for producing the artificial samples range from schemes where jittered copies of each sample are added to the training data, to techniques that use advanced generative models to mimic various degradations or deformations to real or synthetic samples. Data augmentation has had a long successful history in a variety of fields including printed and handwritten text recognition [8], [9], [10], object detection and classification [11], [12], face recognition [13], classification of imbalanced datasets [14], [15], and people and action detection in still images or video [16], [17]. A good example of the success of data augmentation methods for training is reference [16], in which a system trained on deformed samples of only 11 subjects achieves a classification result which is only slightly outperformed by a similar system that was trained on a dataset containing hundreds of real subjects. When the deformed samples were added to the real data, the performance of the system exceeded that of the the previous state-of-the-art system.

In recent years development of software tools that similarly allow researchers to supplement or modify existing medical imaging datasets has been suggested as an alternative to the costly incremental collection of additional real samples. More specifically, these tools are designed to alter the distribution of lesion sizes, shapes, locations, and foreground-background interactions in the original dataset, and thereby potentially improve the training and testing of a CAD system or enable performing perceptual studies which would have been otherwise impossible to conduct due to the limitations of the original dataset. These tools follow two general strategies. In the first approach, artificial lesion shapes that are simulated based on various deformation or cluster growth models are inserted into raw projection data or reconstructed clinical images. Examples of such methods can be found in [18], [19], [20] for lung nodules, in [21], [22] for mammography, and in [23] for digital breast tomosynthesis (DBT). While these techniques can be helpful in assessing general

trends in applications such as validation of nodule volume estimation algorithms or assessment of detection tasks, the use of near spherical shapes and simplistic texture and contrast models is a limiting factor that prevents some of these techniques from fully replicating the wide range of characteristics of real lesions seen in clinical images. A second strategy is to extract lesions from one set of clinical images and insert them into alternate locations in a different set using image processing techniques [24], [25], [26]. The primary advantage of the latter approach is that the selected lesions are natural by design, and therefore the aforementioned limitations of the mathematically simulated lesions are avoided.

In some imaging modalities such as CT, the original projection data is in general not available, and re-projection from reconstructed data is usually complicated and prone to artifacts unless all the acquisition parameters are known and the image formation physics is modeled in great detail. Consequently, forming a composite image using real or simulated lesions is mainly performed in the image domain using a technique referred to as alpha blending. This method involves linearly interpolating the intensities of the source and target images using a weighted sum whose coefficients (weights) are defined by an alpha matte over each of the pixels in the two images [27] (in the context of this paper, the source image is the image that contains a lesion to be blended into a presumably lesion-free area in a target image):

$$f_{alpha}(x, y) = \alpha(x, y)f_s(x, y) + (1 - \alpha(x, y))f_t(x, y), \quad (1)$$

where α is the alpha matte, and f_{alpha} , f_s and f_t designate the blend output, source, and target images, respectively. Since the matte is set manually, the selection of regions of interest in source and target images requires great care so that salient features do not clash with one another. Moreover, hiding the boundary between the inserted object and the surrounding target image is very difficult even when techniques such as feathering (blurring) are used in order to suppress these undesirable artifacts. Overall this is a time consuming process, and the perceived quality of the final composite is largely dependent on the skill of each individual user.

In this study we describe the development of an image based tool that allows for simple selection and seamless insertion of real pulmonary nodules from one real CT exam into another in two steps: the user first identifies a region of interest (ROI) by casually drawing a rectangular region around a nodule, and then selects the center of the desired insertion area in the target slice. The drag and drop capability is achieved by combining the powerful Poisson editing technique, which was originally developed for computational photography [28], with several novel extensions of our own that are designed to facilitate the extraction and insertion of pulmonary nodules. A preliminary version of this work was published in a conference proceedings [29]. Here we extend this previous study with substantial improvements to the algorithms and a more comprehensive evaluation of results.

The main contributions of this paper are summarized as follows. To our knowledge, this is the first time a technique based on Poisson editing has been utilized for medical imaging as

part of a unified approach that addresses issues such as ease of use, limiting impact of user skill, ability to insert non-isolated nodules (e.g. ones connected to vessels), ability to insert a nodule into areas with complex background (e.g. areas with high texture or salient features), and proper transfer of noise properties. In addition, we describe a method to apply different transformations to the shape and contrast of a nodule. Combined with our insertion method, this will allow us to augment a given dataset by altering the distribution of nodule shapes, contrasts, and locations, as well as creating new lesion-background combinations. Finally, we demonstrate the viability of our lesion insertion tool through a reader study using clinical images, as well as objective assessment based on simulated phantoms. This objective assessment analysis is the first time the performance and properties of a lesion insertion technique are being quantified using figures of merit (FOM) based on the area under the ROC curve (AUC) of the Hotelling observer (HO), and the noise power spectrum (NPS). The AUC of the HO was used to assess the effect of the blending process on the detectability of nodules, and the NPS was used to evaluate the similarity of the noise properties of blended nodules compared to native ones.

The remainder of this paper is organized as follows. In Section II we describe the various components of the proposed image blending system which include methods to mask and remove undesirable objects in the source image, reshape the gradient field of nodules connected to other objects, seamlessly blend a source nodule into a target image, and appropriately transfer the noise properties from the target image to the blended image. Examples of performance in clinical data and validation results based on simulated lung phantoms are subsequently presented in Section III before concluding in Section IV.

II. Methods

Gradient domain compositing techniques have found widespread adoption in the computer vision field of computational photography in applications such as image cloning [30], panorama stitching [31], tone management [32], and scene completion [33]. In this section we describe one of the most powerful of such algorithms, Poisson editing [28], and discuss several novel extensions that enables its application to pulmonary nodules in chest CT. We will also present a method to transfer the noise from the target image to the blended result for cases where the noise properties in the original source and target slices do not match.

A. Preprocessing and Blending

Poisson image editing is an effective and versatile technique for seamless image composition. The principal underlying mechanism in this method is the Poisson partial differential equation with Dirichlet boundary conditions. Given the boundary values and the Laplacian of the interior values of a scalar function, we can use the Poisson equation to find a unique solution for the function across the entirety of its domain. Similarly, this process can be used to interpolate the values along the boundary of the target image's insertion area inwards under the guidance of the gradient field of the source image. The result is the seamless blending of the source object into the target image. More formally, let D be the domain under $\mathbb{R}^2$ over which the target image is defined, and let Ω be a closed subset of D with boundary $\partial\Omega$. Using the notation f_s and f_t for the source and target images, respectively,

and f for the desired blend result over the interior of Ω , we can find f by solving the following minimization problem:

$$\min_f \iint_{\Omega} |\nabla f - \nabla f_s|^2; \text{with } f|_{\partial\Omega} = f_t|_{\partial\Omega}, \quad (2)$$

where $\nabla = \left[\frac{\partial}{\partial x} \frac{\partial}{\partial y} \right]$ is the gradient operator. Fig. 1 illustrates the notation used here.

Equivalently, the unique solution to Eq. (2) can be obtained using the Poisson equation with Dirichlet boundary conditions:

$$\Delta f = \text{div} \nabla f_s \text{ over } \Omega, \text{ and } f|_{\partial\Omega} = f_t|_{\partial\Omega}, \quad (3)$$

where $\Delta = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right)$ is the Laplacian operator and $\text{div} = \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} \right)$ denotes the divergence operator. Eq. (3) above can be discretized and solved using iterative numerical methods such as Gauss-Seidel with successive over-relaxation [34] or sparse solvers.

The image composite f obtained by solving Eq. (3) completely replaces the original contents of the target image within Ω with the contents of f_s . If we need to maintain the salient features from the target image within the insertion area, we can modify the guidance field by replacing the gradient field of the source with the gradient field of the target image at spatial locations where the magnitude of the gradient in the target image is higher than that of the source image. By combining the gradient fields of source and target images in this fashion, we allow the salient elements of both images to coexist alongside each other in the composite f :

$$\Delta f = \begin{cases} \text{div} \nabla f_t & \text{if } |\nabla f_t| > |\nabla f_s|, \\ \text{div} \nabla f_s & \text{otherwise.} \end{cases} \quad (4)$$

We can use Eq. (4) to define the guidance field for instance when the target image is highly textured or a vessel is running through the insertion area. The direct solution from Eq. (3) in such cases would result in the lack of texture or abrupt disappearance of the vessel within Ω and thereby provide a visual and anatomical discrepancy. Fig. 2 includes the effect of direct insertion of the source ROI onto the target image in order to demonstrate the significant differences between the contents of the source and target ROIs, and in particular the resulting obvious discontinuities along the borders of the paste ROI. As alpha blending only involves a weighted sum of the source and target ROIs, it should be evident from these examples that hiding those discontinuities using this method would be extremely difficult if not impossible, thereby necessitating additional manual processing steps as mentioned earlier. Fig. 2 also shows examples of using the guidance field from Eq. (3) versus using the guidance field from Eq. (4). In these examples we see that the diffusion of the high intensity

pixels along the edges of the boundary causes severe blurring when using the original guidance field, whereas mixing the gradients allows for the natural continuation of the salient structures within the blend area without any noticeable artifacts along the borders.

Many nodules are connected to the vascular structure or the pleural boundary of the lung and it is necessary to properly identify and remove these attachments prior to image blending. Moreover, the initial boundary drawn by the user might contain extraneous isolated anatomical structures (e.g. a fissure passing through the source ROI) that need not be carried over to the target image. We therefore preprocess f_s with two new extensions to Poisson editing that we refer to as masking and gradient field shaping prior to computing f .

Assuming for simplicity that there is only a single extraneous object within f_s , and referring to the area covered by this object as Ω_{ext} , the simplest way to remove this object is to replace the value of the gradient field of f_s over Ω_{ext} with zeros, and to solve the following equation in order to obtain a modified source image f_s^{mod} :

$$\Delta f_s^{mod} = 0 \text{ over } \Omega_{ext}, \text{ and } f_s^{mod}|_{\partial\Omega_{ext}} = f_s|_{\partial\Omega_{ext}}. \quad (5)$$

However, this solution (which is equivalent to membrane interpolation) produces a seamless but blurry result due to the inward diffusion of pixels along Ω_{ext} . A more suitable alternative is to replace the gradient field of f_s over Ω_{ext} with that of a background region and conduct a guided interpolation for f_s^{mod} as described in Eq. (3) in order to obtain a natural texture within Ω_{ext} . Since a background region with proper size might be difficult to find, we use the Poisson reconstruction mechanism to generate a seamless tile of arbitrary size from a smaller patch, and follow the masking procedure below to remove the undesirable objects:

1. User selects a small rectangular patch of background pixels f_s^{bgnd} in the vicinity of f_s within the source slice.
2. Define a new rectangular image g_s the same size as f_s^{bgnd} as follows:

$$g_s(p) = \begin{cases} f_s^{bgnd}(p) & \text{if } p \in \text{interior,} \\ 0.5 \cdot (f_s^{bgnd}(p) + f_s^{bgnd}(p')) & \text{if } p \in \text{top or bottom,} \\ 0.5 \cdot (f_s^{bgnd}(p) + f_s^{bgnd}(p')) & \text{if } p \in \text{left or right.} \end{cases}$$

where p' is the pixel across from p on the other side of the image (e.g. if p is in the top row, p' is pixel on the same column in the bottom row of the image).

Using ∇f_s^{bgnd} as the guidance field and the values along the borders of g_s as the boundary values, we then apply the same process as Eq. (3) to obtain a tile image g_s^{tile} that can be seamlessly stacked.

3. Stack copies of the individual reconstructed tile g_s^{tile} such that the tiled background is at least as large as f_s . An example demonstrating the seamless tiling of a texture pattern is shown in Fig. 3.

4. Based on a binary segmentation of f_s and given the area covered by the main body of the nodule Ψ_0 , identify all extraneous objects including those connected to the nodule and replace the gradient field corresponding to their locations in f_s with the gradient field of the tiled background at matching locations (see Sec. III-A for information on how Ψ_0 is identified).

Since extraneous objects that are not attached to the main body of the nodule are most often enclosed within the low gradient regions of the interior of the lung, the masking process described above can replicate a continuous smooth texture without visible artifacts. For objects attached to the nodule, we need an additional procedure to shape the gradient field in the vicinity of the nodule segment that is being detached in order to replicate the natural blurring seen across the edges of the rest of the nodule:

1. As show in Fig. 4, let Ψ_0 denote the boundary of the nodule (not to be confused with the casual boundary drawn by the user. Fig. 6 illustrates the distinction between the two). Identify $\partial\Psi_0^{detach}$ as the boundary segment that is immediately adjacent to the area covered by the attachment, and $\overline{\partial\Psi_0^{detach}}$ as the rest of the nodule boundary (i.e. $\overline{\partial\Psi_0^{detach}} = \partial\Psi_0 - \partial\Psi_0^{detach}$).
2. Compute the magnitude of gradients along $\overline{\partial\Psi_0^{detach}}$ and store them in a list $\overline{G_0}$. Set the magnitude of gradients along the pixels on $\partial\Psi_0^{detach}$ equal to the 75th percentile of values stored in $\overline{G_0}$, and the orientations of each gradient vector as being orthogonal to $\partial\Psi_0^{detach}$ at each corresponding pixel.
3. Dilate the nodule surface consecutively using a 3×3 structuring element. Let Ψ_i denote the boundary of the dilated nodule at iteration i , with corresponding $\partial\Psi_i^{detach}$ and $\overline{\partial\Psi_i^{detach}}$. Repeat step 2 to define a new list $\overline{G_i}$ and set the gradient magnitudes and orientations for pixels on $\partial\Psi_i^{detach}$.

In our work, we found that a maximum of three iterations is sufficient for creating a boundary with consistent tapering along the edges. Once both masking and gradient field shaping are completed, we can use the resulting modified gradient field as the guidance field and the values along the boundary of f_s as the boundary values in order to compute f_s^{mod} in a manner similar to that shown in Eq. (3).

Fig. 5 shows an example where the source image contains broth isolated and attached extraneous objects. When gradient field shaping it not used, the low gradients imposed along the boundary of the nodule cause significant diffusion artifacts along the edges. The correction to the gradient field on the other hand avoid these artifacts and creates a natural looking border around the main body of the nodule.

B. Calculation of Optimal Boundary

Poisson editing is prone to causing halos and bleeding artifacts in cases where the boundaries of the source image and target insertion area are not consistent with one another.

To better demonstrate this point, using a change of variables $f' = f - f_s$ we can rewrite Eq. (2) as follows:

$$\min_{f'} \iint_{\Omega} |\nabla f'|^2; \text{with } f'|_{\partial\Omega} = (f_t - f_s)|_{\partial\Omega}, \quad (6)$$

which is equivalent to the following Laplace equation:

$$\Delta f' = 0 \text{ over } \Omega, \text{ and } f'|_{\partial\Omega} = (f_t - f_s)|_{\partial\Omega}. \quad (7)$$

This shows that in order to obtain the blend result f , one could first solve for the membrane interpolant in Eq. (7) and then add f_s to the result. Since f_s does not depend on the solution of the problem, any artifacts seen in the blended image are a function of the differences between source and target images along the insertion boundary.

It can be shown that the energy $\iint_{\Omega} |\nabla f'|^2$ in Eq. (6) will approach zero if and only if all pixels along the boundaries of source and target images satisfy $(f_t - f_s)|_{\Omega} = k$ where k is a constant [35]. In order to avoid interpolation artifacts, we therefore need to choose a boundary which exhibits the most uniform set of differences along its path. However, manually finding and tracing such a boundary is very difficult as both the source and target slices in CT are textured and often contain salient features in the vicinity of the desired ROIs. Here we use an expectation maximization approach to automatically obtain such an optimal boundary Ω_{opt} given the source and target images, and a casually drawn initial boundary (e.g. a rectangular bounding box) for the source object [36].

Let Ω_0 denote the casually drawn initial boundary by the user, and Ψ_0 denote the nodule boundary as before. The goal of optimal boundary estimation is to find Ω_{opt} such that the following cost function is minimized for a yet to be determined constant k :

$$E(\partial\Omega, k) = \sum_{p \in \partial\Omega} (f_t(p) - f_s(p) - k)^2, \text{ s.t. } \Psi_0 \subset \Omega \subset \Omega_0. \quad (8)$$

For a given boundary Ω , the value of k that optimizes Eq. (8) can be found by setting the derivative of E with respect to k to zero:

$$k = \frac{1}{|\partial\Omega|} \sum_{p \in \partial\Omega} (f_t(p) - f_s(p)), \quad (9)$$

where $|\Omega|$ indicates the length of the boundary Ω . Since both k and Ω_{opt} are unknown, the iterative scheme described below is used to compute them:

1. Initialize Ω as Ω_0 .

2. Find current value of k based on Eq. (9).
3. Find the path that minimizes the cost function in Eq. (8) given current value of k (details of this step described further below).
4. Repeat steps 2 and 3 until the energy in Eq. (8) does not decrease anymore.

To find the best path in Step 3 above, we define a graph where each node corresponds to a pixel located on the interior of the original boundary Ω_0 , and the edges between the nodes follow the 4-connectivity pattern of neighboring pixels inside Ω_0 (i.e. an edge exists between two nodes only if their corresponding pixels are 4-neighbors). The cost of traversing along an edge to a node q is then set equal to $(f_\lambda(q) - f_s(q) - k)^2$ so that the overall cost along the nodes of a boundary will conform to the cost function defined earlier (here for simplicity a node and its corresponding pixel are used interchangeably). Based on this definition, the problem of finding the path that minimizes Eq. (8) for a given value of k becomes equivalent to solving for the shortest path through the graph.

The optimal path should form a closed band around Ψ_0 that encompasses the whole nodule without dissecting the main body of the nodule. We therefore further modify the cost matrix of the graph in two stages. First, the cost of traversing along an edge to any node corresponding to a pixel within Ψ_0 is set to infinity. Next, we enforce the path to enclose the nodule by creating a cut through the graph that consists of nodes on a line connecting Ω_0 to Ψ_0 as shown in Fig. 6. We use the nodes situated immediately on one side of the cut as starting points for candidate boundaries, while the nodes located on the cut are candidate ending points. The cost of traversing from any starting node to any end node is then set to infinity. The optimal boundary can now be found by pairing each starting node with a neighboring node located on the cut, and looping over all such combinations. In each iteration, we find the minimum cost boundary for a given pair of starting and end nodes using the Dijkstra algorithm [37] which is a 2D dynamic programming technique. Finally, the path that produces the globally minimum cost is selected as the optimal boundary Ω_{opt} . To reduce the required computation time, the cut should be selected as the one that has the shortest length as it will produce the smallest number of possible pairs of start and end nodes in the loop.

Fig. 7 shows an example of blending a source image into a dense and textured area within a target slice from a different case. The original bounding box contains the pleural boundary of this lung and even though masking and gradient field shaping were used to remove the extraneous areas, the boundary conditions of the target image cause significant visible artifacts in the reconstructed image. The optimal boundary on the other hand avoids areas with large discrepancies relative to the modified source image, and thereby results in a seamless insertion.

C. Noise Transfer

Seamless boundaries might not be sufficient in creating a realistic look if the noise properties of the source and target images differ significantly due to differences in acquisition parameters. Hence we introduce a noise transfer scheme as a final processing step.

Non-local mean [38] is a very effective edge preserving denoising algorithm that utilizes the existing redundancies in an image to perform non-local averaging of pixels with similar neighborhoods. Two main parameters are used to tune the performance of this technique. The first parameter specifies the size of the square similarity neighborhood. The second parameter determines the decay rate of an exponential function that weights the averaging of pixel values according to the degree of similarity between their corresponding neighborhoods. We extract the noise layer by subtracting the non-local means filtered target image from the original target image. The noise layer is subsequently added to the composite image from the previous steps over the insertion area. The edge preserving property of the non-focal means algorithm is very important as we need the noise layer to be devoid of any high frequency components belonging to the actual structure of the target image (e.g. edges of vessels that pass through the insertion region).

Similarly, we can prefilter the source image using the non-local means algorithm in cases where the noise content of the source image is higher than that of the target image. This filtered image can then be used as the input to the lesion blending tool.

III. Performance and Results

Human interpretation of data is a common approach in the evaluation of medical image quality. In addition to providing several examples of the application of our system to clinical images, in this section we report the results of a reader study we conducted with three experienced radiologists using clinical images. The readers were presented with original nodules, as well as their inserted counterparts. In each case, the readers were asked to provide a score indicating their confidence that the nodule being viewed was original (i.e. not inserted). The scores from each reader were then analyzed using ROC methodology to determine the perceived realism of original and inserted clinical nodules.

CT exams produced at different institutions using different scanner hardware, acquisition parameters, and reconstruction algorithms can exhibit large differences in terms of noise characteristics. A mismatch between the properties of the source and target slices in the image blending problem is therefore inevitable in a large number of cases. Assessment of quality using human observers in situations when there are a large number of parameters or protocols governing the performance of a system is very challenging, as many cases generally need to be interpreted for each individual setting. Compared to human readers, mathematical observer models provide a more efficient tool for assessing image quality for certain clinical tasks, and in particular allow us to make a larger number of comparisons to better understand the performance of a given system under various operating conditions. Human observers are also not properly equipped to accurately quantify certain aspects of image quality such as the noise properties of a system, whereas some numerical FOMs are able to provide such characterization. In order to characterize the impact of blending under a number of scenarios where a mismatch between imaging parameters of source and target images exists, we also report the results of two sets of experiments based on lung phantoms simulated using the package described in [39]. The first experiment characterizes the noise properties of blended images and the effect of the noise transfer procedure according to two

numerical FOMs, while the second experiment measures the detectability of blended images using a model observer.

The package used here to simulate helical CT images works in two stages: forward projection followed by image reconstruction. In the first stage, ray tracing is used to calculate the projection data based on parameters that define the scanner geometry (e.g. source-to-isocenter, source-to-detector distances, etc.), scanner system properties (e.g. number of detector rows and elements, pitch, etc.), and object properties (e.g. geometric shape, location, and attenuation coefficient). The package uses an idealized photon-counting detector and a mono-energetic point x-ray source. The forward projection stage incorporates Poisson noise but ignores x-ray scatter in the simulation of the x-ray transport process. The reconstruction stage uses a common filtered back projection (FBP) based algorithm for helical CT [40] and allows for several filter shapes and bandwidths. Finally, the reconstructed slices are recombined along the axial direction using a moving average filter whose width is determined based on the desired slice thickness. The interested reader can refer to [39] for more details.

A. Examples of Performance in Clinical Images

The Lung Image Database Consortium (LIDC) dataset is the largest publicly available CT library of pulmonary nodules [41]. Initiated by the National Cancer Institute (NCI), one of the primary motivations behind creating this library was to facilitate the training and testing of CAD systems by providing researchers with a validated and well characterizes database. The dataset consists of 1018 CT scans from 1010 distinct patients, and each case was reviewed by four experienced thoracic radiologists. The scans were obtained at seven academic institutions, using a wide range of scanner models and manufacturers, and scan protocols (i.e. tube voltage, current, reconstruction kernel, in-plane resolution, etc.). For every 3mm nodule, each radiologist also delineated the boundaries of the lesion on each slice, and provided scores for various subjective characteristics such as malignancy, spiculation, lobulation, subtlety ranking, etc. Overall the database contains 7371 lesions marked as nodule by at least one radiologist, of which 2669 were identified as nodule >3mm by at least one radiologist. Due to the inherent uncertainty and variability of radiologists' assessment of lesions, only 928 out of the 2669 3mm nodules were unanimously marked by all four radiologists. It can be seen that despite the massive effort and resources spent in putting together this dataset, the number of agreed upon 3mm nodules remains relatively small. The ability to supplement or modify this dataset would therefore be very useful in a variety of applications including CAD development and testing.

The examples of clinical images shown thus far and those appearing in this section are taken from the LIDC dataset. In every example and for every nodule, the segmentations provided by each of the radiologists are extracted from the associated XML files, and the union of the delineations is taken as the true nodule segmentation (referred to as Ψ_0 in Sec. II). Attachments and other extraneous objects were identified by first applying Otsu binarization [42] within the user selected initial boundary and then subtracting the area covered by the nodule. They were subsequently removed according to the methods outlined in Sec. II. In datasets other than LIDC where the manual delineation of a nodule might not be available,

Ψ_0 can instead be obtained from a computer segmentation algorithm. It is important to note that the segmentation algorithm need not be very accurate, as Ψ_0 is only required to encompass the main body of the nodule, but not necessarily delineate its precise boundary.

Fig. 8 shows additional examples of the performance of our proposed method in blending consecutive slices of the same nodule from one case onto consecutive slices of a different target case. The user only selected the initial boundary and center of insertion area in a single slice of the source and target cases, respectively, in order to obtain the blend result across all slices. Notice that in each slice the nodule is properly separated from the pleural lung section, and seamlessly inserted in the target images while preserving the vascular structure and texture in the vicinity of the insertion area. Even though the blending process relies on 2D optimal boundaries, the seamless inward guided interpolation of these boundaries across consecutive slices of the target insertion site results in consistent appearance of the inserted nodule when it is observed in the coronal and sagittal views. An example can be seen in Fig. 9 that shows all three orthogonal views of an original nodule and its blended counterpart. Since our blending technique is gradient based, nodules with mixed densities and ground glass opacities are blended into new locations in the same fashion as dense nodules. Fig. 10 shows an example of blending a ground glass nodule into a new location using the same steps as in Fig. 8.

The number of instances of nodules with rare shapes, locations, or background interactions in a particular dataset can be increased by applying various transformations to the shapes and contrasts of the original nodules. Shape transformations can be directly applied to the source image in the image domain, while changes to the nodule contrast profile can be obtained by modifying the guidance field. These transformed samples can subsequently be inserted into the same or different areas of same or different target images in order to create multiple new samples in a single pass given different parameterizations of the desired transformations and each pair of user selected source and target ROIs. Augmenting the original dataset in this way can help improve training of under-represented nodule types and reduce overfitting. Fig. 11 shows examples of blending for a nodule that was modified according to single affine transformations of scaling, rotation, and shear to obtain new shapes, as well as uniform scaling of the guidance field to produce a different nodule contrast. Other transformations such as non-uniform contrast scaling, combinations of affine transformations, and warping are also possible, but not shown here.

Clinical nodules sometimes exhibit strong directionality, or have significant long range interactions with their surrounding tissue. For instance, the growth of a pleural nodule is often inhibited on the side neighboring the pleural boundary. This causes a directional growth toward the inside of the lung. Another example is when a nodule has large supporting structures such as extensive vascular connections. While the series of techniques described in Sec. II simplify the selection of blend ROIs in the source and target images, these methods cannot replicate such distinct anatomical effects within the target image. In such circumstances it is necessary to choose a region in the target slice that best matches the surrounding tissue in the source image as for instance shown in Fig. 12 where a pleural nodule is inserted into the pleura in the target image. When the anatomy of the insertion area within the target image is drastically different from the tissues surrounding the nodule in the

source image, the blend result will look artificial despite having seamless boundaries. An example is shown in Fig. 13 where a nodule with significant directional growth and a large attached vessel is inserted into an empty area in the target image.

B. Reader Study With Clinical CT Scans

We performed a reader study to demonstrate that the appearance of pulmonary nodules inserted into clinical images using our system can be similar to that of the original nodules. The study consisted of nodules taken from the LIDC dataset. In order to demonstrate that the basic principle of our technique is viable, and to simplify the selection of insertion areas, the entire volume containing a nodule was extracted from a CT scan, and then blended into a different location on the same scan. The insertion locations were selected by the first author who does not have medical training. The insertion location was selected such that the entire lesion remained within the lung boundaries after insertion, and the slice numbers containing an original nodule and its blended counterpart did not overlap. The inserted nodule was often placed in a different lobe or the opposite lung. However, as discussed in Sec. III-A, nodules can exhibit directionality if they are close to the pleura or fissures, or if they have significant vascular attachments. In addition, when a large nodule that causes a significant displacement of surrounding vessels or tissue in its original location is inserted at a random location in a target image, it may look unnatural although the insertion may be seamless. In such cases, we selected a region in the target slices with a similar appearance as the surrounding tissue in the source image. The shape or contrast of the lesion were not transformed prior to insertion. A total of 55 original nodules and their inserted counterparts were used for the study, resulting in a total of 110 nodules.

As discussed in Sec. III-A, pleural nodules need to be inserted into similar pleural areas of the lung, and have limited utility in terms of creating dataset variability for data augmentation. Pleural nodules were therefore excluded from the study. Since larger nodules are more likely to be malignant and therefore more relevant for various applications including CAD development, and to stress-test our insertion system with difficult cases, we selected mainly larger nodules with an average principal axis length (PAL) of 10.7 ± 3.1 mm for inclusion in this study. The same method as described in [43] was used to define PAL, where a 3D frame was fitted to the voxels containing each nodule. The largest eigenvector passing through the center of gravity of the voxels was then computed using principal components analysis, and PAL was computed as the distance between the two opposite intersecting points of this eigenvector and the 3D frame. Different radiologists who participated in collection of the the LIDC dataset produced different delineations for each nodule. We computed a PAL based on delineations from each of the radiologists who detected the nodule, and then took the mean of these individual PALs to obtain an average PAL for each nodule. Fig. 14 shows additional information about the nodules used in the study such as the distribution of sizes, average lobulation, sphericity, spiculation, and malignancy scores, and number of slices in each case.

An additional set of 20 nodules (10 original plus 10 inserted counterparts) were used at the beginning of each session of the study for training the readers. During training, the readers

were informed whether each case being viewed contained an original nodule, or an inserted one so that they could develop and refine their scoring strategy.

We applied for and received an exemption from the Institutional Review Board (IRB) for the study. The readers were presented with all the CT slices containing the nodules, as well as two padding slices (one slice before the first appearance of the nodule and one after its last appearance). The padding slices were included to investigate whether the transition from normal lung parenchyma on one slice to a nodule on the neighboring slice was similar for original and inserted nodules. The cases were viewed over two sessions on a DICOM calibrated high resolution NEC PA301W monitor (55 cases per session). Ordering of the cases was randomized for each reader, and the randomization was set up such that an original nodule and its inserted counterpart were not included in the same session. A washout period of at least one week between the two sessions was used to further reduce the chance that a reader could remember the appearance or decision for a nodule previously seen in the first session.

The user interface for the study was set to display whole slices at the default lung window and level (1500HU and -600HU, respectively), but the readers were free to adjust the window and level for each case. Only the relevant slices were available for viewing in each case (slices containing the nodule plus the padding slices). No time limit was imposed on the readers and they were free to scroll through all available slices and adjust the zoom level. A bounding box around each original or inserted nodule was used to mark its location. The size of the bounding box was selected such that it contained the entire nodule, as well as some extra padding so that for inserted nodules the entire insertion area was contained within the bounding box. The readers had the option to switch off the bounding box once they had seen the nodule location so that they could view the area surrounding the nodule without any obstructions. Scoring was based on a continuous scale between 0–100, with 0 indicating definitely inserted, and 100 for definitely original. The readers were not informed about how many original or inserted nodules they would view, and instead only knew the total number of cases.

The three radiologists in our study on average had over 27 years of clinical experience. The scores from each reader were analyzed using ROC methodology. In the ROC analysis, actual positive and actual negative cases were defined as original and inserted nodules, respectively, and the decision variable was the score provided by the radiologist. The goal of our study was to determine whether radiologists can confidently distinguish between original and inserted nodules. Hence AUC values closer to 0.5 (random guessing) would indicate that inserted lesions are very similar in appearance to original lesions. Table I summarizes the study results for the three readers. The average AUC across the readers was 0.63 ± 0.03 , indicating that inserted lesions are difficult to distinguish from original ones.

The user interface for the reader study included a comment box for each case being viewed, and the readers were encouraged to type the reasoning behind their score for a particular case in the provided space. Both the reader comments and the debriefing process that followed the conclusion of the study indicated that the readers' primary scoring strategy was to use higher level reasoning according to the placement of the nodule within the slice and

its interaction with surrounding tissue. For instance, when the consideration pointed out earlier in Sec. III-B regarding vascular connections for larger nodules was not carefully followed, that was used as a clue to score the nodule lower. This was further confirmation that the readers were not finding artifacts introduced by the insertion algorithm, such as gray-level discontinuities or texture differences, but instead relied on anatomical cues. As also pointed out in Sec. III-A, lack of modeling of the biological aspects of growth of a nodule, such as vessel growth and tissue displacement, are issues that affect both image based and physics based lesion insertion methods. Placement issues can be mitigated through the use of improved instructions on proper strategies that consider basic anatomical rules or biological plausibility in selecting an insertion area.

C. Simulated Phantom Experiment: Noise Properties

While the standard deviation of noise is easy to compute, it does not provide information about the correlations present in the noise, i.e. noise texture, and is therefore an incomplete measure. The 2D noise power spectrum (NPS) [44], [45] on the other hand provides such information under the assumption of wide sense stationarity, and in practice is found by estimating the variance of the discrete Fourier transform (DFT) of the noise image at different spatial frequencies for a given ROI:

$$\text{NPS}(u, v) \approx \frac{1}{N} \sum_{j=1}^N |\text{DFT}\{n_j\}(u, v)|^2, \quad (10)$$

where u and v are spatial frequencies in the x and y directions, and n_j is the j^{th} 2D noise image obtained by subtracting the noise free image (in our case estimated as the mean of N noisy realizations) from the j^{th} noisy image realization. The zero-mean noise image represents the stochastic component of the image, i.e. the part that changes every time the experiment is repeated through simulating the acquisition process for the native samples or image blending for the blended samples. The shape of the NPS describes the distribution of noise power in frequency domain, e.g. concentration of noise power in lower frequencies is indicative of a coarse texture in the image domain, whereas a concentration in higher frequencies suggests a fine texture in the image domain. The sum of NPS is equal to the mean variance of pixel values in the ROI.

In the simulations presented in this section, we generated phantoms according to a range of incident photon counts to represent different tube currents, and a range of reconstruction filter bandwidths to represent differences between reconstruction filters for source and target slices. To assess the noise properties, and in particular the effectiveness of the noise transfer step, we measure the NPS of the blended images obtained from different combinations of photon counts and reconstruction filters bandwidths for the source and target slices, and in each case compare them to the NPS computed from realizations generated with acquisition parameters (i.e., photon counts and reconstruction filter bandwidths) matching those of the target images. The degree of similarity between the noise textures of the blended and native images is analyzed based on the 2D correlation of their respective spectra which we will denote as ρ_{NPS} , as defined in Esq. 11 where $\overline{\text{NPS}}$ denotes the mean of the 2D spectrum.

$$\rho_{\text{NPS}} = \frac{\sum_u \sum_v (\text{NPS}_{\text{blend}}(u, v) - \overline{\text{NPS}_{\text{blend}}})(\text{NPS}_{\text{target}}(u, v) - \overline{\text{NPS}_{\text{target}}})}{\sqrt{(\sum_u \sum_v (\text{NPS}_{\text{blend}}(u, v) - \overline{\text{NPS}_{\text{blend}}})^2)(\sum_u \sum_v (\text{NPS}_{\text{target}}(u, v) - \overline{\text{NPS}_{\text{target}}})^2)}}, \quad (11)$$

To determine the agreement between the magnitudes of noise in the blended and native images, we use the ratio between the integrals of the spectra:

$$\phi_{\text{NPS}} = \frac{\sum_u \sum_v \text{NPS}_{\text{blend}}(u, v)}{\sum_u \sum_v \text{NPS}_{\text{target}}(u, v)}. \quad (12)$$

For both ρ_{NPS} and ϕ_{NPS} , a value closer to one indicates a higher similarity between the NPS spectra.

The simulated phantoms in this experiment resembled left and right lungs, and contained a spherical nodule that was 6mm in diameter in the left lung as shown in Fig. 15. The attenuation level of the nodules was chosen to create a visible contrast relative to the surrounding background. The in-plane pixel size in x and y directions and slice interval were all set to 1mm. For image reconstruction, we used a Hanning filter where the cutoff frequency was set at 50% to 100% of Nyquist frequency in steps of 10% (six different bandwidths in total). In the discussion that follows, we will refer to reconstructions from a given filter bandwidth as *hann*** where ** is the corresponding cutoff frequency (e.g. *hann50* has cutoff set at 50% of Nyquist frequency). The photon count of the realizations also varied across a range of values. The blend ROI was a 13×13 region centered at the center of the nodule in the central slice, and optimal boundary option was used in each case to insert the nodule in the same location within the target image. The ROI for calculation of NPS was a 30×30 region centered on the nodule in the central slice that contained it, and therefore captured the noise texture from both the inside and outside of the; blend area for the blended nodules. The number of realizations used to compute the NPS for each comparison was set at 250 ($N = 250$ in Eq. (10)).

The first set of measurements were made for *hann100* reconstructions when noise transfer is not applied, across all simulated combinations of photon counts for source and target images, which we will denote as I_0^{source} and I_0^{target} , respectively. Fig. 16 summarizes these results for ϕ_{NPS} and ρ_{NPS} for each pair of photon counts. As expected, blending from a high I_0^{source} onto a low I_0^{target} results in $\phi_{\text{NPS}} < 1$ since the source image contains less noise than the target, while the opposite combination results in $\phi_{\text{NPS}} > 1$. Moreover, the magnitude of ϕ_{NPS} is directly dependent on the degree of mismatch between the associated photon counts. For cases where the photon counts are matched ϕ_{NPS} is slightly larger than one. The 2D correlation ρ_{NPS} on the other hand is relatively high in most cases, except when blending

from a low I_0^{source} onto a high I_0^{target} . This could be explained by the fact that the target image in such cases does not exhibit a high enough variability within the insertion ROI and the resulting suboptimal boundary conditions cause reconstruction artifacts when solving Eq. (3). A process is thus needed to improve the noise properties of the blend images in cases where the NPS of source and target images differ substantially.

To assess the effectiveness of the noise transfer procedure outlined in Sec. II-C in closing the gap between the noise properties of blended and native images, we conducted a second set of simulations that emulated the circumstance under which the target image has higher noise content relative to the source image. The source photon count I_0^{source} was fixed at $5e+05$ across the six reconstruction filter bandwidths described earlier, and the target images were generated using *hann100* across a range of lower photon counts. For each combination of source and target images, we chose a fixed 3×3 similarity window and varied the decay rate of the nonlocal means algorithm so that the difference between the pixel variances of the target and blended images are minimized. Fig. 17 summarizes the results for ϕ_{NPS} and ρ_{NPS} (the values shown here for blended images with noise transfer are from the experimentally found optimal value of the decay rate). It can be seen that across various combinations of reconstruction filter bandwidths and photon counts for source and target images, the noise transfer procedure leads to an improvement in both ϕ_{NPS} and ρ_{NPS} for the blended images. Figs. 18 and 19 show actual examples of this improvement for the NPS spectra and blend results for one of the combinations of parameters from Fig. 17.

D. Simulated Phantom Experiment: Nodule Detectability

Lesion detection is one of the most common tasks in clinical radiology. Measuring the detectability of lesions is therefore an important aspect in assessing the performance of an imaging system. For our image blending algorithm, it is important to understand how blending affects the detectability of a nodule at a given location in the image. The Hotelling observer (HO) [46] is a numerical model observer which casts the location-known-exactly (LKE) detection problem as a binary classification task, and is known for its effectiveness and correlation with actual human performance for certain tasks across various modalities including CT [47], [48], [49], [50]. In this experiment, we compare the detectability of native and blended nodules using the HO for lung phantoms stimulated using the multi-detector CT (MDCT) simulation software described earlier.

The HO is a linear binary classifier that: computes a template $\vec{\omega}$ from the two classes of data, one with a signal present and the other with the signal absent. Assuming each image is reshaped into a $K \times 1$ column vector where K is the total number of pixels in the image, the template is calculated based on estimates of the means and covariance matrices of the two classes:

$$\vec{\omega} = S^{-1}(\vec{m}_1 - \vec{m}_0) \text{ with } S = \frac{1}{2}(S_1 + S_0), \quad (13)$$

where $\vec{m}_1$ and $\vec{m}_0$ are $K \times 1$ vectors denoting the sample means, and S_1 and S_0 are $K \times K$ matrices denoting the estimated covariance matrices for the signal present and signal absent training samples, respectively. Each test image $\vec{b}$ is projected onto this template to provide a classification score λ :

$$\lambda = \vec{\omega}^t \cdot \vec{b} = \sum_{k=1}^K \omega_k b_k, \quad (14)$$

The scalar λ is subsequently compared against a threshold to determine which class the image belongs to. A common way of assessing the performance of an observer (human or numerical) in a binary classification task is to summarize the performance across all threshold values using the area under the receiver operating characteristic curve (AUC). This FOM has a range between 0.5 for a classifier that is no better than a coin toss, to a 1.0 for perfect classification.

The majority of nodules in lung CT exhibit a high contrast as they are surrounded by air. Slight changes in the LKE detectability of such nodules is of lesser concern since the value of the AUC for the binary detection task described earlier remains unchanged at 1.0. Low contrast nodules however are sensitive to changes in detectability. Here we generated a series of phantoms with the same anatomical structure as those used in Sec. III-C, some with and some without a spherical nodule in the left lung, and under different photon counts to model the effect of different tube currents for the source and target images. The nodules were 4mm in diameter, the in-plane pixel size in x and y directions and slice interval were all set to 1mm. The attenuation coefficient of nodules was chosen to create a low contrast signal in the reconstructed images. A Hanning filter with a cutoff frequency at 100% Nyquist frequency (*hann100*) was used in all cases. For a binary classification task of differentiating ROIs with and without a native nodule, we computed the AUC of the HO over a fixed 11×11 ROI centered at the nodule in the central slice of the volume that contained it. We then repeated this procedure replacing the set of phantoms containing a native nodule with phantoms containing artificially inserted nodules in the central slice. Image blending was done for cases where the photon counts of source and target images matched each other. Using the standard unbiased estimator to find either of the $K \times K$ covariance matrices in Eq. (13), at least K training samples are required so that the resulting matrix will be non-singular [51]. To obtain an accurate estimate however, a rule of thumb is to use 10–100 times K as the number of training samples [52]. In this study we used 2400 realizations for each of the native, blended, and background classes, and split each set such that 2000 samples were used for training (~ 17 times $K = 121$) and the rest for testing the HO at each photon count.

Fig. 20 shows the AUC of the test samples. It can be seen that the detectability of blended nodules at each photon count is somewhat less than that of the native nodules from the same photon count. This difference can be attributed to the fact that even when the optimal boundary option is used, a perfect boundary that satisfies $(f_t - f_s) = k$ along every boundary pixel cannot be generally found. As shown in Section III-C, this results in slightly elevated level of noise in the blended images which can partly explain the reduction in detectability

relative to the background (see the diagonal of ϕ_{NPS} in Fig. 16). To further investigate this point, we also ran an experiment where the source images were first filtered using the non-local means algorithm prior to blending into the target images. The decay rate of this filter was very small, and chosen such that the average of ϕ_{NPS} values computed from the resulting blended images across different photon counts was 1.01 compared to 1.03 for the corresponding average ϕ_{NPS} values from the original blended images. We then computed the AUC values at each photon count as before, and found them to nearly match those of the corresponding native nodules as also shown in Fig. 20.

The results shown in Fig. 20 can be used to identify the appropriate use cases for our lesion blending procedure. According to these results, lesion blending may not be the proper tool to produce additional samples for studies that investigate the effect of system noise on lesion detectability (e.g. perception studies that compare detectability under different image reconstruction algorithms), as any change to lesion detectability would be important in the analysis of study results. On the other hand, lesion blending can be a reasonable approach in situations when matching the detectability of blended and native lesions is not required. Examples of such applications include perceptual experiments where the comparison is between blended lesions on different locations, or training of CAD systems where the main determinant of detectability is anatomical noise.

IV. Conclusion

We have developed a set of techniques for seamless insertion of pulmonary nodules in chest CT images. By incorporating several methods that can automatically handle complex source and target background combinations, we have minimized the amount and thereby impact of user involvement. We have also described a method to transfer the noise properties of the target image to the composite result so that we can apply this method to circumstances where source and target acquisition or reconstruction parameters do not match. In addition, we have provided an objective assessment methodology for our lesion insertion tool, and used it to quantify the impact of our blending procedure on the noise properties and detectability of nodules in simulated phantoms. The results obtained from the reader study using clinical images and the objective assessment using simulated phantom images indicate the viability of using the Poisson editing approach when combined with our proposed extensions for insertion of original or modified lesions in clinical CT images in order to augment and/or otherwise modify the properties of an existing dataset. This general methodology can also be used to insert simulated nodules into clinical images in a more efficient and consistent manner compared to alpha blending. The only parameter in our system is the decay rate of the non-local means filter. While automatic selection of the optimal value for this parameter is feasible in simulated images, this value will need to be selected by the user in clinical images.

Future work is under way to assess the impact of enriching a dataset using the methods described in this paper on training and testing of a CAD system.

Acknowledgments

A.P. was supported in part through a Critical Path grant and an Office of Women's Health grant from the U.S. Food and Drug Administration, and by an appointment to the Research Participation Program at the Center for Devices and Radiological Health administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration. The authors acknowledge the National Cancer Institute and the Foundation for the National Institutes of Health and their critical role in the creation of the free publicly available LIDC/IDRI Database used in this study. The mention of commercial entities, or commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such entities or products by the Department of Health and Human Services or the U.S. Food and Drug Administration.

Biographies

Aria Pezeshk received his B.S degree from Michigan Technological University, Houghton, MI, and his Ph.D degree from the Pennsylvania State University, University Park, PA, both in electrical engineering. From 2011 to 2012, he was an R&D engineer with the Portland Technology Development group at Intel. He is currently an ORISE fellow with the Division of Imaging, Diagnostics, and Software Reliability within the U.S. Food and Drug Administration. His research interests include machine learning, computer vision, and medical imaging.

Berkman Sahiner received his B.S. and M.S. degrees in Electrical Engineering from Middle East Technical University and his Ph.D. in Electrical Engineering from the University of Michigan. He has been a Senior Biomedical Research Scientist at the US Food and Drug Administration and an Adjunct Associate Professor at the University of Michigan since 2009. His research interests include computer-aided diagnosis, image analysis, breast imaging, image perception, and performance assessment methodologies.

Rongping Zeng received the B.S. degree from Tongji University, Shanghai, China in 1997, and the Ph.D. degree in Electrical Engineering from the University of Michigan, Ann Arbor, MI, in 2007. She then joined the Center for Devices and Radiological Health in US Food and Drug Administration as a research scientist within the Office of Science and

Engineering Laboratories. Her research interests include medical imaging physics, image reconstruction, image processing and image quality evaluation.

Adam Wunderlich is a staff fellow at the U.S. Food and Drug Administration, where his research focuses on statistical methodology for task-based image quality assessment. He holds a B.S. in electrical engineering and an M.S. in theoretical and applied mechanics from the University of Illinois at Urbana-Champaign, an M.S. in mathematics from Oregon State University, and a Ph.D. in electrical and computer engineering from the University of Utah

Weijie Chen received his Ph.D in medical physics from the University of Chicago, Chicago, Illinois, in 2007. Since then, he has been a scientist at the Center for Devices and Radiological Health, U.S. Food and Drug Administration (Silver Spring, Maryland). His research interests include statistical assessment methodology for diagnostic devices in general, and ROC methodology and reader studies for medical imaging and computer-aided diagnosis in particular.

Nicholas Petrick is Deputy Director for the Division of Imaging, Diagnostics and Software Reliability within the U.S. Food and Drug Administration, Center for Devices and Radiological Health and is an FDA Senior Biomedical Research Scientist. He earned his B.S. degree from Rochester Institute of Technology in Electrical Engineering and his M.S. and Ph.D degrees from the University of Michigan in Electrical Engineering Systems. His research interest include imaging biomarkers, computer-aided diagnosis, image processing and medical imaging physics.

References

1. Sahiner B, Chan H-P, Hadjiiski LM, Cascade PN, Kazerooni EA, Chughtai AR, Poopat C, Song T, Frank L, Stojanovska J, Attali A. Effect of CAD on radiologists' detection of lung nodules on thoracic CT scans: Analysis of an observer performance study by nodule size. *Academic Radiology*. 2009; 16:1518–1530. [PubMed: 19896069]
2. Zhao Y, Bock G, Vliegenthart R, Klaveren R, Wang Y, Bogoni L, Jong P, Mali W, Ooijen P, Oudkerk M. Performance of computer-aided detection of pulmonary nodules in low-dose CT:

- comparison with double reading by nodule volume. *European Radiology*. 2012; 22(10):2076–2084. [PubMed: 22814824]
3. Das M, Mühlenbruch G, Heinen S, Mahnken AH, Salganicoff M, Stanzel S, Günther RW, Wildberger JE. Performance evaluation of a computer-aided detection algorithm for solid pulmonary nodules in low-dose and standard-dose MDCT chest examinations and its influence on radiologists. *British Journal of Radiology*. 2008; 81(971):841–847. [PubMed: 18941043]
 4. Chan H-P, Sahiner B, Wagner RF, Petrick N. Classifier design for computer-aided diagnosis: Effects of finite sample size on the mean performance of classical and neural network classifiers. *Medical Physics*. 1999; 26(12):2654–2668. [PubMed: 10619251]
 5. Sahiner B, Chan H-P, Petrick N, Wagner RF, Hadjiiski LM. Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size. *Medical Physics*. 2000; 27(7):1509–1522. [PubMed: 10947254]
 6. Raudys S, Jain A. Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1991; 13(3):252–264.
 7. Hanczar B, Hua J, Sima C, Weinstein J, Bittner M, Dougherty ER. Small-sample precision of ROC-related estimates. *Bioinformatics*. Mar; 2010 26(6):822–830. [PubMed: 20130029]
 8. Ha TM, Bunke H. Off-line handwritten numeral recognition by perturbation method. *IEEE Trans Pattern Anal Mach Intell*. May; 1997 19(5):535–539.
 9. Ho T, Baird H. Large-scale simulation studies in image pattern recognition. *IEEE Trans on PAMI*. 1997; 19:1067–1079.
 10. Pezeshk A, Tutwiler RL. Automatic feature extraction and text recognition from scanned topographic maps. *IEEE Trans Geoscience and Remote Sensing*. 2011; 49(12):5047–5063.
 11. Papageorgiou C, Poggio T. A trainable system for object detection. *Int J Comput Vision*. Jun; 2000 38(1):15–33.
 12. Krizhevsky, A., Sutskever, I., Hinton, GE. ImageNet classification with deep convolutional neural networks. In: Bartlett, P.Pereira, F.Burges, C.Bottou, L., Weinberger, K., editors. *Advances in Neural Information Processing Systems (NIPS)*. 2012. p. 1106–1114.
 13. Heisele B, Serre T, Poggio T. A component-based framework for face detection and identification. *Int J Comput Vision*. 2007; 74(2):167–181.
 14. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*. Jun; 2002 16(1):321–357.
 15. Melville, P., Mooney, R. Constructing diverse classifier ensembles using artificial training examples. *Proceedings of the 18th International Joint Conference on Artificial Intelligence*; Acapulco, México. 2003. p. 505–510.
 16. Pishchulin L, Jain A, Wojek C, Andriluka M, Thormhlen T, Schiele B. Learning people detection models from few training samples. *Proc IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)*. 2011:1473–1480.
 17. Chen C-Y, Grauman K. Watching unlabeled video helps learn new human actions from very few labeled snapshots. *Proc IEEE Conference on Computer Vision and Pattern Recognition (CVPR'13)*. 2013:572–579.
 18. Zhang X, Olcott E, Raffy P, Yu N, Chui H. Simulating solid lung nodules in MDCT images for CAD evaluation: modeling, validation, and applications. *Proc SPIE Medical Imaging: Computer-Aided Diagnosis*. Mar.2007
 19. Li X, Samei E, Delong DM, Jones RP, Gaca AM, Hollingsworth CL, Maxfield CM, Carrico CWT, Frush DP. Three-dimensional simulation of lung nodules for paediatric multidetector array CT. *British Journal of Radiology*. 2009; 82(977):401–411. [PubMed: 19153182]
 20. Shin H-O, Blietz M, Frericks B, Baus S, Savellano D, Galanski M. Insertion of virtual pulmonary nodules in CT data of the chest: development of a software tool. *European Radiology*. 2006; 16(11):2567–2574. [PubMed: 16819607]
 21. Saunders R, Samei E, Baker J, Delong D. Simulation of mammographic lesions. *Academic Radiology*. 2006; 13(7):860–870. [PubMed: 16777560]

22. Rashidnasab A, Elangovan P, Yip M, Diaz O, Dance DR, Young KC, Wells K. Simulation and assessment of realistic breast lesions using fractal growth models. *Physics in Medicine and Biology*. 2013; 58(16):5613–5627. [PubMed: 23892735]
23. Vaz MS, Besnehard Q, Marchessoux C. 3D lesion insertion in digital breast tomosynthesis images. *Proc SPIE Medical Imaging: Physics of Medical Imaging*. 2011; 7961:79 615Z1–79 615Z10.
24. Ambrosini RD, O'Dell WG. Realistic simulated lung nodule dataset for testing CAD detection and sizing. *Proc SPIE Medical Imaging: Computer-Aided Diagnosis*. Mar.2010
25. Peskin, AP., Dima, AA. Modeling clinical tumors to create reference data for tumor volume measurement. In: Bebis, G.Boyle, R.Parvin, B.Koracin, D.Chung, R.Hammond, R.Hussain, M.Kar-Han, T.Crawfis, R.Thalmann, D.Kao, D., Avila, L., editors. *Advances in Visual Computing, ser Lecture Notes in Computer Science*. Vol. 6454. Springer; Berlin Heidelberg: 2010. p. 736-746.
26. Madsen MT, Berbaum KS, Scharzt KM, Caldwell RT. Improved implementation of the abnormality manipulation software tools. *Proc SPIE Medical Imaging: Image Perception, Observer Performance, and Technology Assessment*. Mar.2011
27. Porter T, Duff T. Compositing digital images. *SIGGRAPH Comput Graph*. Jan; 1984 18(3):253–259.
28. Pérez P, Gangnet M, Blake A. Poisson image editing. *Proc ACM SIGGRAPH*. 2003:313–318.
29. Pezeshk A, Sahiner B, Zeng R, Wunderlich A, Chen W, Petrick N. Seamless insertion of real pulmonary nodules in chest CT exams. *Proc SPIE Medical Imaging: Computer-Aided Diagnosis*. 2014:90 351K–90 351K-10.
30. Agarwala A, Dontcheva M, Agrawala M, Drucker S, Colburn A, Curless B, Salesin D, Cohen M. Interactive digital photomontage. *ACM Transactions on Graphics*. Aug; 2004 23(3):294–302.
31. Zomet A, Levin A, Peleg S, Weiss Y. Seamless image stitching by minimizing false edges. *IEEE Transactions on Image Processing*. 2006; 15(4):969–977. [PubMed: 16579382]
32. Bae S, Paris S, Durand F. Two-scale tone management for photographic look. *ACM Transactions on Graphics*. Jul; 2006 25(3):637–645.
33. Hays J, Efros AA. Scene completion using millions of photographs. *ACM Transactions on Graphics*. Jul.2007 26(3)
34. Barrett, R., Berry, M., Chan, TF., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., der Vorst, HV. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. 2nd. Philadelphia, PA: SIAM; 1994.
35. Zwillinger, D. *Handbook of Differential Equations*. 3rd. Boston, MA: Academic Press; 1997.
36. Jia J, Sun J, Tang C-K, Shum H-Y. Drag and drop pasting. *Proc ACM SIGGRAPH*. 2006:631–636.
37. Dijkstra E. A note on two problems in connexion with graphs. *Numerische Mathematik*. 1959; 1(1):269–271.
38. Buades A, Coll B, Morel J-M. A non-local algorithm for image denoising. *Proc IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)*. 2005:60–65.
39. Gavrielides MA, Zeng R, Kinnard L, Myers KJ, Petrick N. Information-theoretic approach for analyzing bias and variance in lung nodule size estimation with CT: A phantom study. *IEEE Trans Med Imag*. 2010; 29(10):1795–1807.
40. Hsieh, J. *Computed Tomography - Principles, Design, Artifacts and Recent Advances*. Bellingham, WA: SPIE Press; 2003.
41. Armato SG III, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*. 2011; 38(2):915–931. [PubMed: 21452728]
42. Otsu N. A Threshold Selection Method from Gray-level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*. 1979; 9(1):62–66.
43. Tan J, Pu J, Zheng B, Wang X, Leader JK. Computerized comprehensive data analysis of lung imaging database consortium (LIDC). *Medical Physics*. 2010; 37(7):3802–3808. [PubMed: 20831088]
44. Riederer SJ, Pelc NJ, Chesler DA. The noise power spectrum in computed x-ray tomography. *Physics in Medicine and Biology*. 1978; 23(3):446–454. [PubMed: 674361]
45. Barrett, HH., Myers, KJ. *Foundations of Image Science*. Hoboken, NJ: Wiley-Interscience; 2004.

46. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. Proc National Academy of Science USA. 1993; 90:9758–9765.
47. Wunderlich A, Noo F. Image covariance and lesion detectability in direct fan-beam x-ray computed tomography. Physics in Medicine and Biology. 2008; 53(10):2471–2492. [PubMed: 18424878]
48. Leng S, Yu L, Zhang Y, Carter R, Toledano AY, McCollough CH. Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain. Medical Physics. 2013; 40(8):–.
49. LaRoque SJ, Sidky EY, Edwards DC, Pan X. Evaluation of the channelized Hotelling observer for signal detection in 2D tomographic imaging. Proc SPIE Medical Imaging: Image Perception, Observer Performance, and Technology. 2007; 6515:651 514–651 514-9.
50. Sanchez AA, Sidky EY, Reiser I, Pan X. Comparison of human and Hotelling observer performance for a fan-beam CT signal detection task. Medical Physics. 2013; 40(3):–.
51. Johnson, R., Wichern, D. Applied Multivariate Statistical Analysis. 6th. Pearson Prentice Hall; 2007.
52. Gallas BD, Barrett HH. Validating the use of channels to estimate the ideal linear observer. J Opt Soc Am A. Sep; 2003 20(9):1725–1738.
53. Chen W, Gallas BD, Yousef WA. Classifier variability: Accounting for training and testing. Pattern Recognition. Jul; 2012 45(7):2661–2671.

Fig. 1.

Poisson blending: f is the inward interpolation of f_t given the boundary conditions in Ω and the guidance field ∇f_s derived from the source image f_s

Fig. 2.

Gradient mixing; In each row from left to right: original source image, original target image from a different case, result of direct insertion of source ROI onto target, Poisson blend results without gradient mixing, Poisson blend results with gradient mixing. The fissure in the first row and the pleural boundary in the second row are properly handled when gradient mixing is used. Orange and red boxes show the original boundary drawn by the user

Fig. 3.
Example of tiling on a 2×2 grid; Left: regular stacking of a texture pattern. Right: seamless stacking of the same texture pattern.

Fig. 4.

Gradient field shaping: dashed lines show the boundaries of the dilated nodule surface, and the shaded pattern shows the area covered by extraneous attachment

Fig. 5.

Effect of gradient field shaping on the boundary of the nodule. From left to right: Original source image, masking result without gradient field shaping, masking result with gradient field shaping. Images are shown in false color to improve the visibility of diffusion artifacts along both the upper and lower boundaries of the nodule in the middle image.

Fig. 6.
Optimal boundary: Ω_{opt} should be located between Ω_0 and Ψ_0 .

Fig. 7.

Effect of optimal boundary; From left to right: original source image, original target image from a different case, blend results using; original boundary (notice the darkening artifact in the center of image), blend results using optimal boundary with both boundaries superimposed, blend results using optimal boundary without boundaries superimposed. Orange and red boxes show the original boundary drawn by the user, and the optimal boundary is shown in yellow.

Fig. 8.

Performance in LIDC: blending consecutive slices of a nodule onto a different case. In each row from left to right: source image, target image, blend results using optimal boundary with both boundaries superimposed, blend results using optimal boundary without boundaries superimposed. Orange and red boxes show the initial boundary drawn by the user, and the optimal boundary is shown in yellow.

Fig. 9.

Comparison of orthogonal views of an original nodule (first row) compared to its inserted counterpart (second row). In each row from left to right: coronal view, axial view, sagittal view. Crosshairs mark the location of the nodule in each view.

Fig. 10.

Blending a ground glass nodule (average radiologist texture score of 1.3). From left to right: source image, target image, blend result using optimal boundary with both boundaries superimposed, blend result using optimal boundary without boundaries superimposed. Orange and red boxes show the initial boundary drawn by the user, and the optimal boundary is shown in yellow.

Fig. 11.

Augmenting a dataset with transformed copies of a nodule. From left to right: blending the original, contrast scaled (0.8 first row, 1.4 second row), size scaled (0.85 both rows), rotated (30° both rows), and sheared (0.25 horizontal both rows) copies of a nodule into the same target image.

Fig. 12.

Example of successful blending of a pleural nodule into the pleura of a different case. From left to right: source image, target image, blend result. Red boxes show the initial boundary drawn by the user.

Fig. 13.

Example of limitations of proposed system. When the nodule has significant supporting anatomical structures or directional growth, seamless blending will not be sufficient in creating a natural looking result if the anatomy of the insertion area does not match that of the source image. From left to right: source image, target image, blend result. Red boxes show the original boundary drawn by the user.

Fig. 14.

Histograms of various properties of the clinical nodules included in the reader study. For each nodule, malignancy, lobulation, sphericity, and spiculation scores from all LIDC radiologists were pooled to produce its average score (in each corresponding histogram, scores range between 1–5, and a higher score indicates increased perceived prominence of the observed property). First row from left to right: Histograms of average malignancy (2.77 ± 0.75), average lobulation (1.59 ± 0.69), and average sphericity (3.91 ± 0.59) scores. Second row from left to right: histograms of average spiculation (1.37 ± 0.47) scores, average principal axis lengths ($10.74 \pm 3.15\text{mm}$), and number of slices including padding (6.15 ± 2.05). Numbers in parenthesis indicate mean and standard deviation of values for each histogram.

Fig. 15. Phantom shape used in simulations. The attenuation coefficient of the nodule has been increased in this picture in order to improve visibility.

Fig. 16.

Comparison of noise properties of blended and native images from *hann100* across different combinations of I_0^{source} and I_0^{target} when noise transfer is not used. Left: ϕ_{NPS} (deeper shades of blue and red indicate smaller and larger ratios of noise magnitude, respectively). Right: ρ_{NPS} (darker shades indicate higher correlation). The tables for ϕ_{NPS} and ρ_{NPS} are color coded differently to better highlight the variability of values in each of them.

Fig. 17.

Improvement of noise properties of blended images with noise transfer over those without noise transfer. Source images are taken from different reconstruction filter bandwidths with $I_0^{\text{source}} = 5e+05$, and target images are taken from *hann100* reconstructions at different photon counts. Error bars were obtained from 1000 bootstrap samples, and show the 95% confidence intervals for the case of *hann70* source images (confidence intervals for other reconstruction filter bandwidths were similar in size but not shown in order to avoid congestion in the plots). Left: ϕ_{NPS} . Right: ρ_{NPS} .

Fig. 18.

Matching the NPS of native images using noise transfer for one of the combination of source and target images shown in Fig. 17; Source images taken from *hann60* with $I_0^{source}=5e+05$, target taken from *hann100* with $I_0^{target}=2e+04$. From left to right: NPS of native images, NPS of original blend images, and NPS of blended images after noise transfer. Images are shown in same scale and using false color mapping to improve visibility.

Fig. 19.

Actual realizations from same combination of acquisition and blend parameters as in Fig 18. From left to right: source, target, original blend result, noise extracted from target, and blended image after noise transfer. Images are shown in the same scale.

Fig. 20.

Model observer study with simulated phantoms: comparison of test AUC of native nodules with blended nodules with matching source and target photon counts. Error bars show the 95% confidence intervals. AUC values and corresponding confidence intervals computed using the methods described in [53].

Table I

Summary of reader study results.

	Years of clinical experience	AUC $\pm$ standard deviation
Reader1	> 30	0.52 $\pm$ 0.06
Reader2	> 35	0.65 $\pm$ 0.05
Reader3	> 17	0.71 $\pm$ 0.05
Average	> 27	0.63 $\pm$ 0.03