

# Article Applying Self-Supervised Learning to Image Quality Assessment in Chest CT Imaging

Eléonore Pouget <sup>1,2,\*</sup> and Véronique Dedieu <sup>1,2</sup>

- <sup>1</sup> Department of Medical Physics, Jean Perrin Comprehensive Cancer Center, F-63000 Clermont-Ferrand, France; veronique.dedieu@clermont.unicancer.fr
- <sup>2</sup> UMR 1240 INSERM IMoST, University of Clermont-Ferrand, F-63000 Clermont-Ferrand, France
- \* Correspondence: eleonore.pouget@clermont.unicancer.fr

**Abstract:** Many new reconstruction techniques have been deployed to allow low-dose CT examinations. Such reconstruction techniques exhibit nonlinear properties, which strengthen the need for a task-based measure of image quality. The Hotelling observer (HO) is the optimal linear observer and provides a lower bound of the Bayesian ideal observer detection performance. However, its computational complexity impedes its widespread practical usage. To address this issue, we proposed a self-supervised learning (SSL)-based model observer to provide accurate estimates of HO performance in very low-dose chest CT images. Our approach involved a two-stage model combining a convolutional denoising auto-encoder (CDAE) for feature extraction and dimensionality reduction and a support vector machine for classification. To evaluate this approach, we conducted signal detection tasks employing chest CT images with different noise structures generated by computer-based simulations. We compared this approach with two supervised learning-based methods: a single-layer neural network (SLNN) and a convolutional neural network (CNN). The results showed that the CDAE-based model was able to achieve similar detection performance to the HO. In addition, it outperformed both SLNN and CNN when a reduced number of training images was considered. The proposed approach holds promise for optimizing low-dose CT protocols across scanner platforms.

**Keywords:** self-supervised learning; feature representation learning; convolutional denoising autoencoder; task-based approach; model observer; chest CT image

# 1. Introduction

Many new computed tomography (CT) reconstruction techniques have been developed in recent years to address the concern of reducing radiation exposure levels [1]. These methods typically exhibit nonlinear properties, which can result in an unfamiliar image texture in very low-dose CT [2]. As traditional physical metrics may not be adequate to assess image quality in nonlinear reconstructions, a task-based paradigm assessment has been advocated [3,4]. For signal detection tasks, the Bayesian ideal observer is optimal but its computation is often analytically intractable [5,6]. As an alternative, the Hotelling observer (HO) is commonly employed. The HO is the ideal linear observer in the sense that it maximizes the signal–noise ratio of the test statistic [6]. Its definition requires inverting the covariance matrix, for which its computational cost makes the HO difficult to implement in clinical practice [7,8].

Recent efforts have demonstrated the ability of supervised learning-based approaches to produce accurate estimates of HO performance [3,6,9–11]. For example, in a preliminary work, Zhou et al. [6] investigated the use of simple linear single-layer neural networks (SLNNs) to approximate the HO test statistic for binary signal detection tasks. They showed that this learning-based method was able to produce accurate estimates of the HO performance. More recently, Kim et al. [3] have demonstrated the feasibility of using convolutional neural networks (CNNs) to provide similar detection performance to the



Citation: Pouget, E.; Dedieu, V. Applying Self-Supervised Learning to Image Quality Assessment in Chest CT Imaging. *Bioengineering* **2024**, *11*, 335. https://doi.org/10.3390/ bioengineering11040335

Academic Editors: Zhiming Luo and Sheng Lian

Received: 9 January 2024 Revised: 25 March 2024 Accepted: 27 March 2024 Published: 29 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). HO in breast CT images. The ability of CNNs to learn sophisticated image representations directly from the image data makes CNNs powerful tools in medical image analysis [12,13]. The accuracy and robustness of such a supervised approach, however, are dependent on the availability of large-scale annotated images [10,13,14]. Such labelled datasets are difficult to obtain in the medical domain due to the complexity of manual annotation and inter- and intra-observer variability [13,14].

In response to this challenge, various methods have been proposed. One approach relies on the concept of transfer learning, whereby general features are extracted from different domains in which large labeled datasets are available [12–14]. However, this approach is often unable to extract discriminative features for medical image analysis problems [12–14]. Another approach is based on self-supervised learning (SSL), which plays a growing role in the field of machine learning due to its capability of learning representations that are transferable to different downstream tasks [15,16]. This learning paradigm reduces the reliance on labeled data by training a model to extract meaningful representations of the input data with no manual labeling required [16]. Recent studies have leveraged the potential of SSL for various tasks such as image segmentation and classification [15–17]. The success of this approach strongly depends on how the pretext tasks are chosen [15]. Among the pretext tasks proposed in the literature, the category of pixel-to-pixel follows the paradigm of generating low-dimensional representations of inputs for use in downstream tasks [15].

In this context, autoencoders (AEs) and extensions such as Sparse AE [18], Denoising AE (DAE) [19,20], or Convolutional AE (CAE) have gained wide popularity because of their ability to extract discriminative and robust features [12,15,21–24]. The use of deep learning techniques in a preprocessing step allows shallow machine learning algorithms, such as support vector machine (SVM), to be used to interpret encoded features for classification [21,22]. This framework has been successfully applied in various domain applications [21,22,25,26]. In particular, CAEs have performed well in medical image classification [12,13].

This study proposes an SSL-based model observer to predict HO performances in very low-dose chest CT images. To the best of our knowledge, this is the first study to apply the SSL framework for designing a mathematical model observer. We hypothesize that integrating the SSL paradigm into model observer pipelines holds promise for implementing the task-based optimization of CT systems. Our approach combined CAE for the feature representation and SVM for the classification task. We compared this approach to state-of-the-art supervised methods through different signal detection tasks of varying difficulty. Computer-based simulations were conducted by addressing different noise structures to investigate the generalization property of the proposed SSL-based model observer. This property is a key factor in ensuring equivalent image quality across a fleet of scanners with increasing differences between reconstruction techniques.

### 2. Materials and Methods

# 2.1. Data

The dataset used in this work comprised images generated from the Lungman multipurpose chest phantom that was reconstructed with various regularization terms proposed for low-dose CT [27]. This dataset was constructed from computer-based simulations, as described in a previous work [28]. The latter showed that the generalization performance of an observer strongly depends on the ability of the feature extractor to filter the noise out and reveal the structures of the original data. We proposed here an extension of this work by investigating the benefit of using deep learning techniques as feature extractors in various noise structures.

To evaluate the generalization properties of the proposed SSL-based model observer, we conducted signal detection tasks over the whole dataset regardless of the regularization strategy in order to be representative of the practical use of the numerical observers [29].

To get as close as possible to clinical tasks, all simulations addressed backgroundknown-statistically (BKS) signal detection tasks [5]. The image data were generated by



**Figure 1.** CT images of the Lungman phantom using (**a**) the Green and (**b**) the quadratic potential function into the regularization term of the iterative reconstruction method employed.

#### 2.1.1. Signal Known Exactly (SKE)

The first signal detection task employed a nodule signal profile given by the following [30,31]:

$$f_{s} = \begin{cases} A_{s} \left[ 1 - \left(\frac{|r_{m} - r_{c}|}{R}\right)^{2} \right]^{z} , |r_{m} - r_{c}| \leq R \\ 0 , otherwise \end{cases}$$
(1)

where  $|r_m - r_c|$  is the distance between the location of the *m*th pixel and the signal center, *R* is the signal radius, and *A*<sub>s</sub> is the parameter that controls the signal amplitude. The parameters were set as follows: *R* was fixed at 4 mm, *z* was fixed at 4, and signal amplitude was set to -870 HU, mimicking the attenuation of low-attenuation ground-glass opacities (GGO) [32]. In this task, the signal was invariant, referring to the signal-known-exactly (SKE) paradigm.

### 2.1.2. Signal Known Statistically (SKS)

For the second signal detection task, we modeled an elliptical profile given by the following:

$$f_s = A_s \exp\left(-(R_{\Phi}(r_m - r_c))^T D^{-1}(R_{\Phi}(r_m - r_c))\right),$$
(2)

where *D* is a diagonal matrix that controls the width of the signal along each direction  $\sigma_x$  and  $\sigma_y$ , and  $R_\theta$  is the Euclidian rotation matrix that controls the angle of rotation for the signal,  $\Phi$  [5,33]. In this study, we fixed  $\sigma_x = 5$  and  $\sigma_y = 1.5$ , and  $\Phi$  was randomly sampled from the set {0°, 45°, 90°, 135°} according to the detection task previously used by Granstedt et al. [5].

### 2.2. Hotelling Observer

The test statistic of a linear model observer, *t*, is computed by the following [3,5,6]:

$$(f) = w^T f, (3)$$

where f is a column vector of the image, and w is the linear observer template [3]. The HO employs the population equivalent of the Fisher linear discriminant, which can be defined as follows [5,6]:

t

$$w_{HO} = \left[\frac{1}{2}(K_0 + K_1)\right]^{-1} \Delta \overline{f},\tag{4}$$

where  $K_i$  is the covariance matrix of the measured data under each hypothesis, and  $\Delta f$  is the mean difference between signal-present and signal-absent images [5,31]. The dimension of the covariance matrix is  $M^2$  where M is the number of image pixels [3]. An accurate

estimate of *K* requires the collect of at least ten times  $M^2$  sample images which makes the HO estimation difficult if not impossible in practical usage [7].

In Equation (4), the inverse covariance matrix serves to decorrelate the noise in the image [31]. This point is particularly interesting to understand to what extent human observers are able to adapt to the statistical properties of the images managed by the reconstruction process, as previously stated by Abbey et al. [31].

### 2.3. Supervised Learning-Based Model Observer

Numerous works have successfully employed supervised learning-based technologies for implementing model observers [3,6,29,34,35]. In this work, a CNN and an SLNN were explored for binary signal detection tasks. The area under the receiver-operating-characteristic curve (AUC) was used for assessing the detection performance. The results obtained were compared to those produced by the proposed SSL-based model observer.

We optimized the structure of the CNN using brute-force searching in the parameter space defined by the depth (from 2 to 10) and the number of filters (4, 8, and 16). The filter size in each convolutional layer was set as  $3 \times 3$  pixels in order to reduce the number of model parameters, as evoked by Kim et al. [3]. No downscaling layer was included to prevent the removal of any high-frequency component [3]. Each convolutional layer was followed by a LeakyReLU activation function to add nonlinear property into the network [3], while a sigmoid function was used after the fully connected layer. The output of the model can thus be interpreted as the posterior probability that an observation belongs to one of the two possible classes in a binary classification problem. In this work, the SLNN was not employed as a linear model and included the sigmoid function as the activation function.

The network training was performed in Tensorflow using a mini-batch stochastic gradient descent algorithm, i.e., the Adam algorithm [5,6,36,37], with 50 epochs to minimize the cross-entropy loss function. During the network training, the train–validation–test scheme [6] was used. We generated 5000 images to prepare the training dataset. Both the validation and testing datasets comprised 200 image pairs. The CNN resulting in the minimum validation cross-entropy was defined as the optimal model in the explored search space [6]. Once the network structure was optimized, the performance of the selected CNN was evaluated as a function of the training dataset size. This point is of particular interest as it represents the principal limitation of implementing CNNs in practical usage.

# 2.4. SSL-Based Model Observer

# 2.4.1. AE

An AE is a specialized type of artificial neural network that consists of two components: the encoder and decoder. The first maps the input data vector X into a hidden representation h, while the latter aims to reconstruct X from the embedding h. This process can be formulated as follows (considering an AE with one hidden layer):

$$h = s_1(W_1X + b_1), X' = s_2(W_2h + b_2)$$
(5)

where  $s_1$  and  $s_2$  are activation functions. The AE is parametrized by the weight matrices  $W_1$  and  $W_2$  and bias vectors  $b_1$  and  $b_2$ . Training an AE involves finding parameters  $\theta = (W_1, W_2, b_1, b_2)$  since they minimize the reconstruction error on a given dataset [23,24]. The cross-entropy was used as the loss function as follows:

$$L_{c}(W_{1}, W_{2}, b_{1}, b_{2}; X) = -\sum_{i=1}^{N} \left[ X_{i} \log \left( X'_{i} \right) + (1 - X_{i}) \log(1 - X'_{i}) \right]$$
(6)

It is worthwhile noting that considering the mean-squared error as the loss function, the optimal linear one-layer AE projects the data onto a subspace defined by its principal

directions [5]. Therefore, the principal component analysis (PCA) method can be thought of as a simplified form of AE. The PCA was performed in this work for comparison.

A weight-decay regularization term can be added to Equation (6) to limit the increase in weights that can occur during the training process, leading to model overfitting [22,23]. Thus, the objective function can be formulated as follows:

1

$$L = L_c + \lambda W^2, \tag{7}$$

where  $W^2$  guarantees the weight matrix having small elements. In this formulation, the penalty term is a quadratic constraint referring to l<sub>2</sub>-norm regularization. The hyperparameter  $\lambda$  controls the strength of the regularization [23]. Classically, this parameter lies in the range [0, 0.1]. We varied the parameter  $\lambda$  over this interval to study its impact on the AE discriminative ability.

Another way to solve the overfitting issue is to consider DAE, which introduces noise in the original data. The DAE is trained to reconstruct the corrupted input, minimizing the loss function between the reconstructed X' and the original data X [23]. The corrupting operation was achieved by randomly setting some pixels in the input sample to zero, leading to more robust features than the traditional AE [19,20]. Compared to the other variants of the AE that ignore local connections between image content, convolutional AEs combine the advantage of a CNN and an AE to preserve spatial information, potentially limiting the redundancy in the parameters [13,22]. They have performed well in image denoising and classification [13], justifying their use in this work.

The structure of a convolutional DAE (CDAE) is similar to that of a DAE, except that it deals with convolutional layers rather than fully connected layers [38,39]. We optimized the structure of the network by performing a Bayesian search [38] in the parameter space defined by the depth (from 1 to 8), the number of filters (4, 8, or 16), and the number of latent units (from 2 to 32). Each convolutional layer comprised filters of  $3 \times 3$  pixels and was followed by a LeakyReLU activation function to add nonlinear properties into the network. A sigmoid function was used after the last convolutional layer of the decoder. The output of the model was thus bounded on the interval [0, 1], which was of interest because the images were expressed in attenuation coefficients  $(cm^{-1})$ . Zero padding was applied in order to maintain the size of the input. Compared to the traditional grid search method, which treats hyper-parameter sets independently, Bayesian optimization is an informed search method that learns from previous iterations [40-42]. As a result, this method can determine optimal hyper-parameters with fewer trials. This is of particular interest for optimizing network architecture where the complexity of grid search grows exponentially with the addition of new parameters. The Bayesian search was performed 20 times on 5000 images extracted from the whole dataset. For each trial, the training and testing sets were generated by a resampling procedure. During network training, the cross-entropy loss function was used, as well as the Adam optimizer [36]. An early-stopping strategy was employed, acting as the hyper-parameter for the number of epochs.

# 2.4.2. SVM

For all learned features, the SVM was used for the classification task. This learning technique, based on a differentiable hinge loss, became popular due to its strong generalization capabilities. To achieve a non-linear separation between classes, a kernel function was introduced in the SVM formulation, including linear, polynomial, sigmoid, and RBF kernels [21]. The RBF function was employed in this study because of its ability to produce complex decision boundaries [40]. In this formulation, two hyper-parameters *C* and  $\sigma$  have to be adjusted. Parameter *C* controls the trade-off between correct classification on training samples and the maximization of the decision function's margin, while  $\sigma$  defines the kernel width [43]. For example, the value of  $\sigma$  being too low makes the radius of influence of the support vector only include the support vector itself, which can lead to overfitting [43]. On the other hand, large values of  $\sigma$  can make the model too constrained [43]. The SVM performance strongly depends on the proper selection of these hyperparameters. To do so, we used a grid search method by applying a five-fold cross-validation (CV) procedure. This strategy was justified by the fact that the search space was limited to only two parameters, *C* and  $\sigma$ . The AUC was used as the score metric for evaluating the quality of the model predictions. Compared to other metrics, the AUC does not require an optimizing threshold for each label. The selection of the optimal values of *C* and  $\sigma$  was carried out on 1000 images extracted from the image database that did not overlap with the images used for the CDAE network structure optimization. After the optimal model was selected, the performance of the SSL-based model observer was assessed as a function of the training dataset size.

#### 3. Results

### 3.1. Optimal Structure of the CNN

The train–validation–test scheme was repeated twenty times to improve the statistical reliability of the results. The validation cross-entropy, plotted in Figure 2, increased with the depth of the CNN for both tasks investigated. This result was most likely due to the overfitting issue that can occur when adding many hidden layers in the neural network.



**Figure 2.** Effect of the convolutional layer number (from two to ten layers) on the CNN performance evaluated during network structure optimization. The error bars displayed correspond to the standard deviation obtained during this process. (a) Values of the validation cross-entropy (left) and AUC (right) for the SKE/BKS detection task; (b) values of the validation cross-entropy (left) and AUC (right) for the SKS/BKS detection task.

When the SKE/BKS detection task was considered, the detection performance (i.e., AUC) evaluated on the testing dataset was 0.96 and 0.89 for the CNN with two and ten convolutional layers, respectively. These values were 0.98 and 0.82 for the SKS/BKS detection task. Therefore, the CNN that possessed two convolutional layers (each having

AUC (SKE/BKS, SKS/BKS)

eight filters) was selected for each detection task considered. It can be noted that the standard deviations associated with the results produced by the CNN increased with the convolutional layer number.

### 3.2. Optimal Parameters for the SSL-Based Model Observer

For both detection tasks, the optimal encoder derived from the Bayesian search comprised one convolutional layer with 16 filters, followed by a LeakyReLu activation function and finally a fully connected layer that processed the features maps into 25 latent units.

In SSL, the regularization hyper-parameter is one of the main parameters influencing the classification accuracy [21]. Table 1 shows how the hyper-parameter  $\lambda$ , which controls the relative importance of the weight-decay regularization applied during the CDAE training process, impacts the choice of *C* and  $\sigma$  in the SVM formulation.

CDAE PCA  $\lambda = 0.00001$  $\lambda = 0.0001$  $\lambda = 0.001$  $\lambda = 0.01$  $10^4, 10^3$  $10^3, 10^2$  $10^1, 10^2$  $10^7, 10^6$  $10^3, 10^2$ C(SKE/BKS, SKS/BKS)  $\sigma(SKE/BKS, SKS/BKS)$ 100, 100 10, 10 1,1 0.01, 0.1 0.1, 0.1

0.77, 0.70

**Table 1.** Optimal values of *C* and  $\sigma$ , and the corresponding AUC for each detection task considered.

It was observed that, as the value of  $\lambda$  increased, the classifier's decision function favored low values of  $\sigma$ , translating interactions between similar observations. This can be explained by the fact that, for high values of  $\lambda$ , the weights learned converge towards zero. As a consequence, the encoder resembled a linear model, leading to underfitting situations. Conversely, low values of  $\lambda$  resulted in unregularized models, leading to the overfitting problem. Therefore,  $\sigma$  acts as a good structural regularizer for the SVM model.

0.88, 0.91

0.63, 0.65

0.78, 0.86

From these results, we fixed the regularization parameter as  $\lambda = 0.001$  for both the detection tasks considered. In this case, the optimal models are positioned on the diagonal of *C* and  $1/\sigma$ . Taking intermediate values of  $\sigma$  and lowering the value of *C* seemed to be a good trade-off between generalization and prediction accuracy. However, lower *C* values generally relate to more support vectors, which can lead to a growing storage requirement and computational complexity, while high *C* values may increase the fitting time [21,43]. In this work, the model seemed to perform equally regardless of the value of *C* for intermediate values of  $\sigma$ . This reflected the fact that no more training points were located within the margin, and the parameter  $\sigma$  was acting alone as a good regularizer [43].

### 3.3. Detectability Comparison between Observers

0.76, 0.74

In these experiments, the detection performance of the selected models was evaluated on a reduced number of training images in order to be consistent with the realizations achievable in clinical practice. Figure 3 shows the AUC and fitting times obtained by the different observers investigated as a function of the training dataset size. We can note that a straight line was plotted to represent the HO performance, as it was estimated from a fixed training dataset.

In both tasks, the CDAE-based model observer provided nearly similar performance to the HO when using at least 1000 images. For the SKE/BKS task, which was more complex, this observer provided higher AUC values than the PCA-based model. This result demonstrated that the CDAE was able to extract more discriminative features for classification than the linear PCA. The CNN and SLNN provided higher detection performance than the HO when a sufficient amount of data was considered (i.e., at least 2500 images). It is worth mentioning that the CDAE-based model outperformed the supervised learning-based model observers on a reduced number of training images (i.e., less than 1000).



**Figure 3.** Effect of the training dataset size on detection performance and fitting times for the supervised learning-based model observers and SSL-based model observers investigated. (**a**) AUC (**left**) and fitting times (**right**) of the selected models for the SKE/BKS detection task; (**b**) AUC (**left**) and fitting times (**right**) of the selected models for the SKS/BKS detection task.

The storage requirement and computational complexity of the CNN were larger than those of the other models investigated. We can attribute this result to the number of variables estimated by the CNN, which was larger than that of the SLNN or SVM.

### 4. Discussion

The current state of SSL research has already achieved promising results in different domains of medicine such as digital pathology or computer-aided diagnosis [15–17,26]. In particular, the SSL paradigm can take advantage of the large volumes of unlabeled datasets available in medical imaging [6,15]. Thus, this approach may address the challenges impeding deep-learning models to become widespread in medical image analysis. Taking inspiration from the literature, this work adopted a two-stage model, which combined the CDAE network for feature extraction and SVM for the classification task, to produce accurate estimates of HO performance in low-dose chest CT images.

The proposed SSL-based model provided similar detection performance to the HO under various noise structures. This point is of interest because the ability of an observer to generalize well over different reconstruction algorithms is crucial to optimizing and standardizing a fleet of scanners [29,44]. We conjecture that the application of a regularization constraint during the CDAE training process gave this model a better generalization capability than that of the linear PCA. Although not presented in this paper, no direct correlation was observed between the AE reconstruction loss and the classification performance. This emphasizes the fact that the features learned by AE do not necessarily guarantee strong discriminative ability [22]. To address this concern, an alternative was proposed by

several authors, which tends to simultaneously minimize the reconstruction loss of AE and the structural risk term of the classifier [22,25]. Such a framework imposes the encoder to project input data into discriminative subspaces for classification and remains a topic of future investigation [22].

The effectiveness of the model strongly depends on the SVM hyper-parameter selection. The results obtained demonstrated that the hyper-parameters have to be tuned with regard to the features learned in the pre-training step. The radius of the RBF kernel  $\sigma$  acted as a good regularizer regardless of the feature extraction technique experimented. Intermediate values of  $\sigma$  seemed suitable in order to prevent overfitting issue on one hand and avoid too "smooth" models on the other hand [43]. For these values of  $\sigma$ , the CDAE-based model was found to be unsensitive to changes in the values of the *C* parameter. Therefore, using small values of *C* allowed a decrease in the training and testing time, which is an important factor to consider for straightforward implementation.

Moreover, the proposed SSL-based model observer provided a higher detection performance than the state-of-the-art supervised learning-based methods when a reduced number of training images was considered. This result is consistent with previous published works, where the authors showed that a large amount of labeled data is required for implementing the CNN [3,6]. In this work, the addition of convolutional layers in the CNN architecture resulted in a decrease in the detection performance. An early-stopping strategy could be employed during the network training to reduce this overfitting problem. Alternatively, a multiple classifier that makes interim decisions instead of one at the end of the network would be used [45].

This study has several limitations. First, our experiments have focused on classification tasks in chest CT images. This work should be extended to other anatomical localizations and diagnostic tasks in order to demonstrate the generalizability of this method as a surrogate for the HO. Second, we intend to extend this work to handle the 3D regularization along both axial and longitudinal directions used by current reconstruction algorithms in CT images. Finally, we will evaluate the hierarchical feature extractor proposed by Ahn et al. [12,13], which introduced a CAE placed atop a pre-trained CNN, to further improve the feature representation of image data.

# 5. Conclusions

In summary, this study proposed an approach that combined deep and shallow learning techniques to match the performance of the HO for low-dose chest CT images. This approach leveraged a large amount of unlabeled data, which can be accessed through archives or shared databases, to learn meaningful representations for enhancing classification and generalization performance. This approach thus reduced the reliance on labeled data. Therefore, the proposed CDAE-based model observer may be easier to deploy than the HO for use as a task-based measure of image quality. We anticipate that the SSL-based approach can contribute to more accurate image quality analysis and can be a useful tool for objective comparison between different CT scanners or reconstruction algorithms.

**Author Contributions:** Conceptualization, E.P.; methodology, E.P.; software, E.P.; validation, E.P and V.D.; formal analysis, E.P.; investigation, E.P.; resources, E.P.; data curation, E.P.; writing—original draft preparation, E.P.; writing—review and editing, E.P. and V.D.; visualization, E.P.; supervision, E.P.; project administration, V.D.; funding acquisition, V.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available on request from the authors.

**Conflicts of Interest:** The authors declare no conflicts of interest.

# References

- 1. Solomon, J.; Lyu, P.; Marin, D.; Samei, E. Noise and spatial resolution properties of a commercially available deep learning-based CT reconstruction algorithm. *Med. Phys.* **2020**, *47*, 3961–3971. [CrossRef] [PubMed]
- Funama, Y.; Nakaura, T.; Hasegawa, A.; Sakabe, D.; Oda, S.; Kidoh, M.; Nagayama, Y.; Hirai, T. Noise power spectrum properties of deep learning-based reconstruction and iterative reconstruction algorithms: Phantom and clinical study. *Eur. J. Radiol.* 2023, 165, 110914. [CrossRef]
- 3. Kim, G.; Han, M.; Shim, H.; Baek, J. A convolutional neural network-based model observer for breast CT images. *Med. Phys.* 2020, 47, 1619–1632. [CrossRef] [PubMed]
- 4. Zhang, Y.; Leng, S.; Yu, L.; Carter, R.E.; McCollough, C.H. Correlation between human and model observer performance for discrimination task in CT. *Phys. Med. Biol.* **2014**, *59*, 3389–3404. [CrossRef] [PubMed]
- 5. Granstedt, J.L.; Zhou, W.; Anastasio, M.A. Approximating the Hotelling Observer with Autoencoder-Learned Efficient Channels for Binary Signal Detection Tasks. *arXiv* 2020, arXiv:2003.02321. [CrossRef] [PubMed]
- 6. Zhou, W.; Li, H.; Anastasio, M.A. Approximating the Ideal Observer and Hotelling Observer for Binary Signal Detection Tasks by Use of Supervised Learning Methods. *IEEE Trans. Med. Imaging* **2019**, *38*, 2456–2468. [CrossRef] [PubMed]
- Gallas, B.D.; Barrett, H.H. Validating the use of channels to estimate the ideal linear observer. J. Opt. Soc. Am. A Opt. Image Sci. Vis. 2003, 20, 1725–1738. [CrossRef] [PubMed]
- 8. Barrett, H.H.; Denny, J.L.; Wagner, R.F.; Myers, K.J. Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance. J. Opt. Soc. Am. A Opt. Image Sci. Vis. 1995, 12, 834–852. [CrossRef] [PubMed]
- 9. Kopp, F.K.; Catalano, M.; Pfeiffer, D.; Fingerle, A.A.; Rummeny, E.J.; Noël, P.B. CNN as model observer in a liver lesion detection task for x-ray computed tomography: A phantom study. *Med. Phys.* **2018**, *45*, 4439–4447. [CrossRef]
- Zhou, W.; Li, H.; Anastasio, M.A. Learning the Hotelling observer for SKE detection tasks by use of supervised learning methods. In *Medical Imaging 2019: Image Perception, Observer Performance, and Technology Assessment*; SPIE: San Diego, CA, USA, 2019; Volume 10952, pp. 41–46. [CrossRef]
- Zhou, W.; Anastasio, M.A. Learning the ideal observer for SKE detection tasks by use of convolutional neural networks (Cum Laude Poster Award). In *Medical Imaging 2018: Image Perception, Observer Performance, and Technology Assessment*; SPIE: San Diego, CA, USA, 2018; Volume 10577, pp. 287–292. [CrossRef]
- Ahn, E.; Kumar, A.; Fulham, M.; Feng, D.D.F.; Kim, J. Unsupervised Domain Adaptation to Classify Medical Images Using Zero-Bias Convolutional Auto-Encoders and Context-Based Feature Augmentation. *IEEE Trans. Med. Imaging* 2020, 39, 2385–2394. [CrossRef]
- Ahn, E.; Kumar, A.; Feng, D.; Fulham, M.; Kim, J. Unsupervised Deep Transfer Feature Learning for Medical Image Classification. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 1915–1918. [CrossRef]
- 14. Ahn, E.; Kim, J.; Kumar, A.; Fulham, M.; Feng, D. Convolutional Sparse Kernel Network for Unsupervised Medical Image Analysis. *Med. Image Anal.* 2019, *56*, 140–151. [CrossRef] [PubMed]
- 15. Chowdhury, A.; Rosenthal, J.; Waring, J.; Umeton, R. Applying Self-Supervised Learning to Medicine: Review of the State of the Art and Medical Implementations. *Informatics* **2021**, *8*, 59. [CrossRef]
- 16. Kwak, M.G.; Su, Y.; Chen, K.; Weidman, D.; Wu, T.; Lure, F.; Li, J. Self-Supervised Contrastive Learning to Predict Alzheimer's Disease Progression with 3D Amyloid-PET. *medRxiv* 2023. [CrossRef]
- 17. Xing, X.; Liang, G.; Wang, C.; Jacobs, N.; Lin, A.L. Self-Supervised Learning Application on COVID-19 Chest X-ray Image Classification Using Masked AutoEncoder. *Bioengineering* **2023**, *10*, 901. [CrossRef] [PubMed]
- 18. Luo, Y.; Wan, Y. A novel efficient method for training sparse auto-encoders. In Proceedings of the 2013 6th International Congress on Image and Signal Processing (CISP), Hangzhou, China, 16–18 December 2013; Volume 2, pp. 1019–1023. [CrossRef]
- Vincent, P.; Larochelle, H.; Bengio, Y.; Manzagol, P.A. Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine Learning, ICML '08. Association for Computing Machinery, Montreal, QC, Canada, 11–15 April 2016; pp. 1096–1103. [CrossRef]
- 20. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
- 21. Al-Qatf, M.; Lasheng, Y.; Al-Habib, M.; Al-Sabahi, K. Deep Learning Approach Combining Sparse Autoencoder With SVM for Network Intrusion Detection. *IEEE Access* 2018, *6*, 52843–52856. [CrossRef]
- 22. Binbusayyis, A.; Vaiyapuri, T. Unsupervised deep learning approach for network intrusion detection combining convolutional autoencoder and one-class SVM. *Appl Intell.* **2021**, *51*, 7094–7108. [CrossRef]
- 23. Meng, Q.; Catchpoole, D.; Skillicom, D.; Kennedy, P.J. Relational autoencoder for feature extraction. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 364–371. [CrossRef]
- 24. Meng, L.; Ding, S.; Xue, Y. Research on denoising sparse autoencoder. Int. J. Mach. Learn. Cyber. 2017, 8, 1719–1729. [CrossRef]
- 25. Liu, Y.; Zhou, S.; Wu, H.; Han, W.; Li, C.; Chen, H. Joint optimization of autoencoder and Self-Supervised Classifier: Anomaly detection of strawberries using hyperspectral imaging. *Comput. Electron. Agric.* **2022**, *198*, 107007. [CrossRef]
- 26. Riquelme, D.; Akhloufi, M.A. Deep Learning for Lung Cancer Nodules Detection and Classification in CT Scans. *AI* **2020**, *1*, 28–67. [CrossRef]

- 27. Multipurpose Chest Phantom N1 "LUNGMAN" | KYOTO KAGAKU. Available online: https://www.kyotokagaku.com/en/products\_data/ph-1\_01/ (accessed on 2 August 2021).
- 28. Pouget, E.; Dedieu, V. Comparison of supervised-learning approaches for designing a channelized observer for image quality assessment in CT. *Med. Phys.* **2023**, *50*, 4282–4295. [CrossRef]
- Brankov, J.G.; Yang, Y.; Wei, L.; El Naqa, I.; Wernick, M.N. Learning a Channelized Observer for Image Quality Assessment. *IEEE Trans. Med. Imaging* 2009, 28, 991–999. [CrossRef]
- 30. Burgess, A.E.; Li, X.; Abbey, C.K. Nodule detection in two-component noise: Toward patient structure. In *Medical Imaging 1997: Image Perception;* SPIE: Newport Beach, CA, USA, 1997; Volume 3036, pp. 2–13. [CrossRef]
- Abbey, C.K.; Barrett, H.H. Human- and model-observer performance in ramp-spectrum noise: Effects of regularization and object variability. J. Opt. Soc. Am. A Opt. Image Sci. Vis. 2001, 18, 473–488. [CrossRef]
- 32. Mathieu, K.B.; Ai, H.; Fox, P.S.; Godoy, M.C.B.; Munden, R.F.; de Groot, P.M.; Pan, T. Radiation dose reduction for CT lung cancer screening using ASIR and MBIR: A phantom study. *J. Appl. Clin. Med. Phys.* **2014**, *15*, 4515. [CrossRef]
- Park, S.; Witten, J.M.; Myers, K.J. Singular Vectors of a Linear Imaging System as Efficient Channels for the Bayesian Ideal Observer. *IEEE Trans. Med. Imaging* 2009, 28, 657–668. [CrossRef]
- Kupinski, M.A.; Edwards, D.C.; Giger, M.L.; Metz, C.E. Ideal observer approximation using Bayesian classification neural networks. *IEEE Trans. Med. Imaging* 2001, 20, 886–899. [CrossRef]
- 35. Massanes, F.; Brankov, J.G. Evaluation of CNN as anthropomorphic model observer. *Med. Imaging* **2017**, *10136*, 101360Q. [CrossRef]
- 36. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. arXiv 2017. [CrossRef]
- 37. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. {TensorFlow}: A System for {Large-Scale} Machine Learning. In Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; pp. 265–283. Available online: https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi (accessed on 21 December 2023).
- 38. Nishio, M.; Nagashima, C.; Hirabayashi, S.; Ohnishi, A.; Sasaki, K.; Sagawa, T.; Hamada, M.; Yamashita, T. Convolutional auto-encoder for image denoising of ultra-low-dose CT. *Heliyon* **2017**, *3*, e00393. [CrossRef]
- Ye, Q.; Liu, C.; Alonso-Betanzos, A. An Unsupervised Deep Feature Learning Model Based on Parallel Convolutional Autoencoder for Intelligent Fault Diagnosis of Main Reducer. *Intell. Neurosci.* 2021, 2021, 8922656. [CrossRef] [PubMed]
- 40. Laref, R.; Losson, E.; Sava, A.; Siadat, M. On the optimization of the support vector machine regression hyperparameters setting for gas sensors array applications. *Chemom. Intell. Lab. Syst.* **2019**, *184*, 22–27. [CrossRef]
- 41. Wu, J.; Chen, X.Y.; Zhang, H.; Xiong, L.D.; Lei, H.; Deng, S.H. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26–40. [CrossRef]
- Nair, A. Grid Search VS Random Search VS Bayesian Optimization. Medium. Published 2 May 2022. Available online: https://towardsdatascience.com/grid-search-vs-random-search-vs-bayesian-optimization-2e68f57c3c46 (accessed on 21 June 2022).
- 43. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- Favazza, C.P.; Duan, X.; Zhang, Y.; Yu, L.; Leng, S.; Kofler, J.M.; Bruesewitz, M.R.; McCollough, C.H. A cross-platform survey of CT image quality and dose from routine abdomen protocols and a method to systematically standardize image quality. *Phys. Med. Biol.* 2015, *60*, 8381–8397. [CrossRef]
- Chan, K.H.; Im, S.K.; Ke, W. A Multiple Classifier Approach for Concatenate-Designed Neural Networks. *Neural Comput Applic.* 2022, 34, 1359–1372. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.