# X-Ray Tomographic Non-Destructive Testing of Manufactured Components using Bimodal Convolutional Deep Belief Network for Sinogram Enhancement

**By Emilien Valat , Thomas Blumensath , Katayoun Farrahi**
University of Southampton, Southampton, SO17 1BJ, UK

## Abstract

When manufacturing high value components, non-destructive imaging techniques are often used to guarantee component integrity and manufacturing standards. X-ray computed tomography (XCT) is a powerful, non-destructive imaging technique that uses X-rays to generate 3D images of the internal structure of an object. However, when imaging objects with complex external geometries or objects manufactured from materials that significantly absorb X-rays, then current techniques often do not provide good images, as not all of the required measurements can be taken. This then means that the computational inverse problem that computes the 3D image from the X-ray data then becomes significantly ill posed.

However, in many cases, prior knowledge about a manufactured object is available. In this research, we want to harness information from Computer Assisted Design (CAD) data to enhance CT acquisitions by generating missing measurements. By doing so, one can hope to get either an image with a better resolution from a normally sampled acquisition or to match the existing resolution from an under-sampled acquisition.

Inspired by Lee et al (2009), we have developed a convolutional variant of the bimodal Deep-Belief Network (DBN) proposed by Ngiam et al (2011) that can be used to estimate missing information. Our architecture allows unsupervised feature extraction from a CT scan and its associated CAD data. In addition, it allows joint learning of the extracted high-level features using an auto-encoder like supervised fine-tuning step. The fine-tuning focuses on interpolating evenly spaced missing acquisitions and extrapolating several consecutive acquisitions at random locations in the sinogram. The training is performed using transfer learning on data of increasing complexity.

We compare the performance achieved by our bimodal convolutional DBN to other machine learning method (Li et al, 2019; Tovey et al, 2019). In addition, we study the improvements in image quality compared to reconstructions where the CAD data is not used.

Our approach to sinogram inpainting is novel in several ways. Firstly, the use of prior knowledge about the objects CAD features is novel. Learning the joint distribution of an object and its CAD allows enhancement of both modalities and with fully connected neural networks, one can also generate the CAD drawing given the scan data and vice-versa. Secondly, the network we develop is novel. To our knowledge, no previously published approach uses the generative power of the energy-based models for sinogram inpainting. In addition, the use of a convolutional layers in a bimodal DBN setting has not previously been explored.

## 1. Introduction

Non destructive testing is a group of analysis techniques used to evaluate the properties of a material without causing damage to it. X-ray Computed Tomography (XCT) can be used for non-destructive testing as it allows the generation of the material's sections by combining X-ray measurements taken from various angles. This technique produces grey-scale images that depict the X-ray attenuation properties of each component of the material, but XCT suffers from several limitations. Among them, complex geometries or opaque-to-X-rays components might degrade the quality of the reconstructed image as the scanning process would produce scarce measurements. On the other hand, high-resolved images require long scan times and hence high X-ray doses that might damage the scanned object.

Prior information about the inspected object is often available. Whether in industrial testing or medical imaging, the Computer Assisted Design (CAD) data or anatomic drawings are often known. Hence, our goal is to examine the use of CAD data to overcome the limitations in imaging complex objects. To tackle the problem of integrating prior-knowledge into XCT, we have chosen to start by investigating the generative power of Deep-Belief Networks (Hinton & Salakhutdinov, 2006). These networks have also been used for multi-modal learning (Ngiam, 2011) but never, to our knowledge, in a setting where the size of the data required the use of convolutional Restricted Boltzmann Machines (Lee, 2009).

## 2. Background : Restricted Boltzmann Machines as Building Blocks

### 2.1. *Restricted Boltzmann Machines*

A RBM is an undirected graphical model with visible units $\mathbf{v}$ and hidden units $\mathbf{h}$. Visible units and hidden units are connected via symmetric connections (W) and have biases, denoted as $\mathbf{a}$ and $\mathbf{b}$, respectively. For a set of parameter $\theta = \{\mathbf{a}, \mathbf{b}, \mathbf{W}\}$ and any given $\mathbf{v}$ and $\mathbf{h}$, an energy function can be defined as

$$E(\mathbf{v}, \mathbf{h}, \theta) = -\mathbf{a}^T\mathbf{v} - \mathbf{b}^T\mathbf{h} - \mathbf{v}^T\mathbf{W}\mathbf{h}. \tag{2.1}$$

This energy function can be used to define a joint probability distribution over the hidden and visible variables $\mathbf{h}$ and $\mathbf{v}$

$$p(\mathbf{v}, \mathbf{h}, \theta) = \frac{e^{-E(\mathbf{v},\mathbf{h},\theta)}}{Z(\theta)}, \tag{2.2}$$

where $Z(\theta)$ is the normalising constant. For binary units, the state of a unit is sampled from

$$p(h_j = 1|\mathbf{v}) = \sigma(b_j + \sum_{j=1}^{M} w_{ij}v_j) \text{ and } p(v_i = 1|\mathbf{h}) = \sigma(a_i + \sum_{j=1}^{M} w_{ij}h_j). \tag{2.3}$$

Binary units are good to model the CAD modality as it can describe the boundaries of a component. The measurement modality requires a model that describe real values.

### 2.2. *Gaussian RBM*

Gaussian RBM have real-valued visible states $\mathbf{v}$ whilst $\mathbf{h}$ are still binary stochastic units. The energy is then defined as

$$E(\mathbf{v}, \mathbf{h}, \theta) = -\sum_i \frac{(a_i - v_i)^2}{2\sigma_i} - \sum_j h_j b_j - \sum_{ij} \frac{v_i}{\sigma_i} W_{ij}^{vh} h_j, \tag{2.4}$$

where the parameter vector $\theta = \{\mathbf{a}, \mathbf{b}, \mathbf{W}, \sigma\}$ now also includes the variance. The state of the visible units is sampled from

$$p(v_i|\mathbf{h}, \theta) = \mathcal{N}(a_i + \sum_{j=1}^{M} \frac{w_{ij}}{\sigma_i} h_j, \sigma_i^2). \tag{2.5}$$

In most cases, as $\sigma$ can be tricky to learn, the input data is renormalised to have 0 mean and unit variance.

### 2.3. *Convolutional RBM*

Fully connected RBMs are good models for small data, but they do not scale well to larger images or large volumetric images as found in real XCT applications. We thus chose to use convolutional RBMs. The hidden layer is now composed of K groups of detection layers with weight-sharing and K groups of pooling layers. A convolutional RBM uses Probabilistic Max-Pooling, an operation that shrinks detection layer representations and ensures at most one unit in a given small area (block) of the detection layer is on.

For all models presented, the exact maximum likelihood is intractable. An approximation of the gradient, the Contrastive Divergence (CD) objective, is hence used in practice during training (Hinton, 2002). In this procedure, a Monte-Carlo Markov Chain is ran to generate samples from the distribution modelled by the RBM.

### 2.4. *Deep Belief Network*

A RBM is defined as 2-layer system. However, to model more complex data and to extract modality-independent representations, increasing the number of layers is necessary. This is done by training deep autoencoders in a greedy, layer wise manner : a RBM is trained and its posteriors used by the next layer as training data.

### 2.5. *Joint layer*

Once high-level features of each modalities have been extracted, they are used to train a joint layer with energy function defined as

$$E(\mathbf{v}, \mathbf{h}, \theta) = -\mathbf{a}_1^T \mathbf{v}_1 - \mathbf{a}_2^T \mathbf{v}_2 - \mathbf{b}^T \mathbf{h} - \mathbf{v}_1^T \mathbf{W}_1 \mathbf{h} - \mathbf{v}_2^T \mathbf{W}_2 \mathbf{h}. \tag{2.6}$$

This joint layer is supposed to capture cross-correlations between the two modalities. It is also trained with CD and the input data are the posteriors of the top RBMs of each modalities : all of the units have a binary state.

The training procedure is composed of several steps. First, for each modality, a feature extraction DBN is trained in a greedy, layer-wise manner. Then, the joint layer is trained with the posteriors of the top RBMs as inputs. Finally, the network is unrolled to build an autoencoder. The goal of this procedure is to find close-to-optimum parameters in an unsupervised fashion and fine-tune them in a supervised manner. The network forms a Bimodal-Deep-Belief-Autoencoder.

## 3. Experiments

### 3.1. *Research goals*

We are interested in answering key questions of the approach, knowing:

- How does the bimodal design performs against simpler designs (RBM and DBN).
- How does the supervised fine-tuning affects the quality of the reconstructions.
- How does the convolutional-RBM architectures scale to large data image.

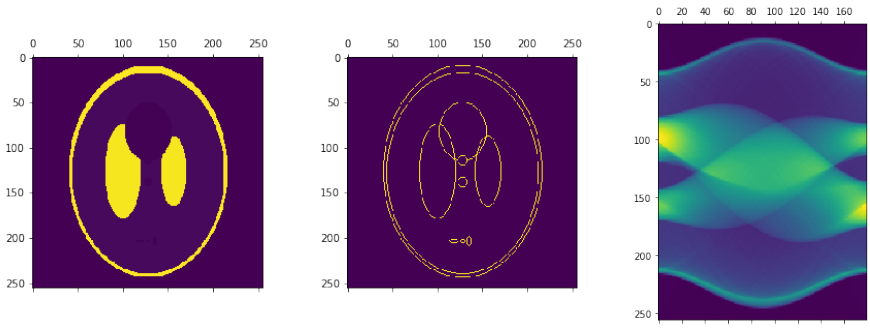


Figure 1: From left to right : The Shepp-Logan phantom, the associated CAD data and the corresponding sinogram

To address these questions, we will first design a database that includes X-ray-like measurements and CAD data. Then, for each architecture, fully-connected or convolutional, we will compare bimodal designs and simpler ones. To quantify the effect of fine-tuning on the networks, we will compare the quality of the image inpainted by a finely-tuned and a non-finely tuned one. Finally, to know how well does the convolutional architecture performs, we will inpaint an acquisition of 256x220 pixels acquisition coming from a 256x256 image, the Shepp-Logan phantom (Shepp & Logan, 1974).

### 3.2. *Datasets*

To test how the multimodal architecture performs at generating missing measurements, our training data will be composed of the CAD data (the boundaries of an object) and the acquisitions : the sinograms which contain the X-ray projection data.

For a given acquisition angle, X-ray beams penetrate the object and their intensity is measured after exit by a detector. An increase in attenuation along the X-ray path will lead to a decrease in X-ray intensity. We focus on parallel beams in a 2D plane through the object, that is, each angle of acquisition produces a 1D vector of intensities. The sinogram is the collection of these 1D projections for different angles. As is standard in x-ray imaging, sinogram intensity values are converted to average X-ray attenuation values along the path, so that large values indicate large attenuation.

We have two datasets :

- a dataset composed of 64x64 pixel images for both modalities.
- a dataset composed of 256x256 pixels images for CAD modality and 256x220 pixels for the sinogram.

The training images will typically come from synthetic data of ellipsoids with random centres and radiuses. From each image, a sinogram and CAD data are generated. To simulate missing acquisitions, a number of randomly selected frames are removed from the sinograms. We chose to use this kind of dataset as it is widely used in the field (Jin et al, 2016; Kelly et al, 2017; Adler et al, 2017).

### 3.3. *Initial experiments*

Training the models on smaller size data first has several advantages. First, the benchmarking of the algorithms is easier as the training is faster. Second, it offers a good insight into the performance of each model. Finally, we have found that the convolutional RBM were easier to train on smaller images that had the same properties as larger ones and that the properties learned could be transferred to the larger images. For all tests, the

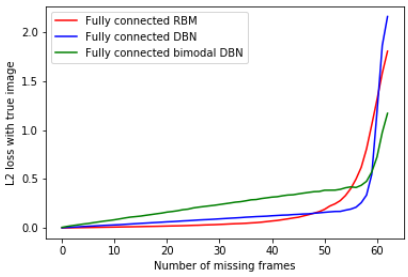



Figure 2: L2-loss between reconstructed image and target image against number of missing frames for fully-connected architectures.

images will consist of one ellipsoid and the associated sinogram. To measure the quality of the reconstructions, we replace the missing frames by the ones generated by the model and then measure the L2-loss between the target sinogram and the reconstructed one.

**The Bimodal-DBAE against single modalities**

In this experiment, we compare the quality of inpainting of a fully-connected RBM, a DBN and a multimodal DBN. The RBM, the DBN and the sinogram modality of the multimodal DBN are trained on the sinogram dataset while the CAD modality of the multimodal DBN is trained on the associated CAD dataset. To measure the quality of inpainting, columns in a given sinogram are replaced by zero vectors : this operation corresponds to a missing acquisition at a given frame. Then, the visible units of each network are clamped with the degraded sinogram. Afterwards, one step of MCMC is ran in each network to generate their representation of the data. Finally, the degraded sinograms's missing frames are replaced by the ones generated by each model. The figure 2 displays, for a given number of missing frames in the degraded sinogram, the L2 loss between the reconstructed image and the target sinogram.

In figure 2, we show that using information from the CAD data provides a better reconstruction for higher numbers of missing frames but at the expense of a worst reconstruction at a lower number of missing frames. One must also point out that a lower error does not always come with a qualitative improvement : figure 3 shows how the reconstruction can get tricky to understand for higher number of missing frames.

Moreover, we found that training the joint RBM was complicated, and as the training of the bimodal RBM relies on this specific layer it is important to consider it as a bottleneck for the whole architecture.

**Impact of fine-tuning on the different architectures** We are interested in seeing how the different architectures are affected by the fine tuning step. In particular, does the second modality help to improve the supervised training?

To answer this question, we "unrolled" each network presented hereupon and realised supervised fine-tuning by minimizing the L2-loss between the reconstructed sinograms and the input one for single modalities networks (RBM and DBN) and the L2-loss between the reconstructed sinograms and CADs for the multimodal DBN.

As one can see in figure 4, the fine-tuning only improves the performance of the reconstruction when two modalities are available. However, in 4d, one can see that the multimodal architecture does not perform better than the traditional DBN, when it comes to generating measurements on a 64x64 sinogram.

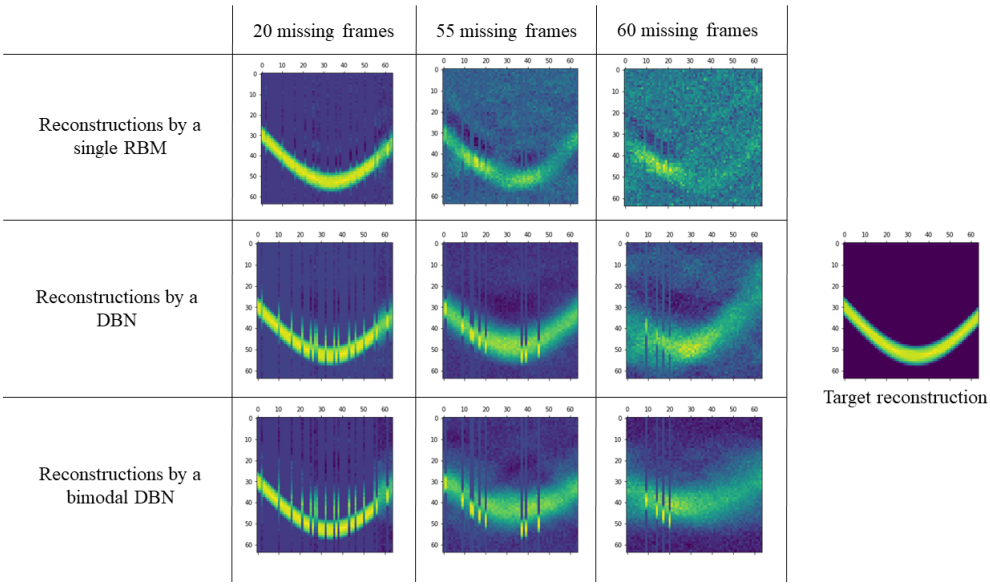
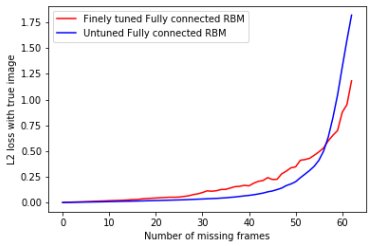
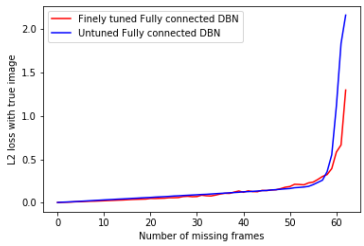


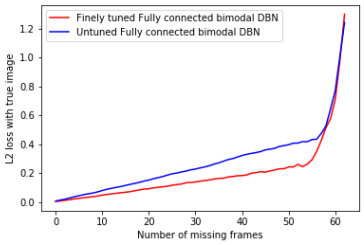

Figure 3: Reconstructions for various numbers of missing frames



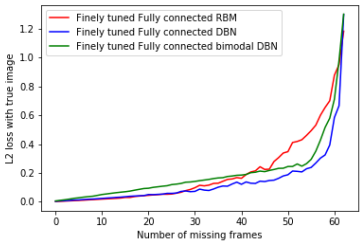

(a) Effect of fine-tuning on RBM




(b) Effect of fine-tuning on DBN




(c) Effect of fine-tuning on multimodal DBN




(d) Fine tuning performance comparison

Figure 4: Comparison of the effect of fine tuning on the different fully connected architectures

### 3.4. *Performances on the Shepp-Logan Phantom : scaling to 256*256 images*

We are now interested in realistically-sized images : 256*220 pixel sinograms along with 256x256 pixel CAD. We chose to implement four layers for the feature extraction of each modality and we have trained the two first layers of each modality on 64x64 images : we found out that they required no further training on the larger images and we were

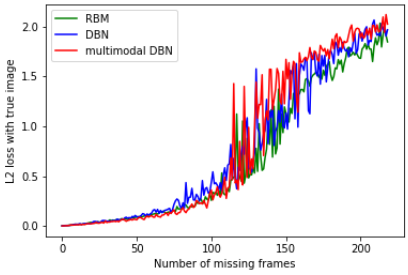



Figure 5: L2-loss between reconstructed image and target Shepp-Logan phantom against number of missing frames for convolutionnal architectures.

satisfied by the reconstructions. The architecture we use can be found in appendix 1. To measure the quality of inpainting, we proceeded the same way as we did for 64x64 images.

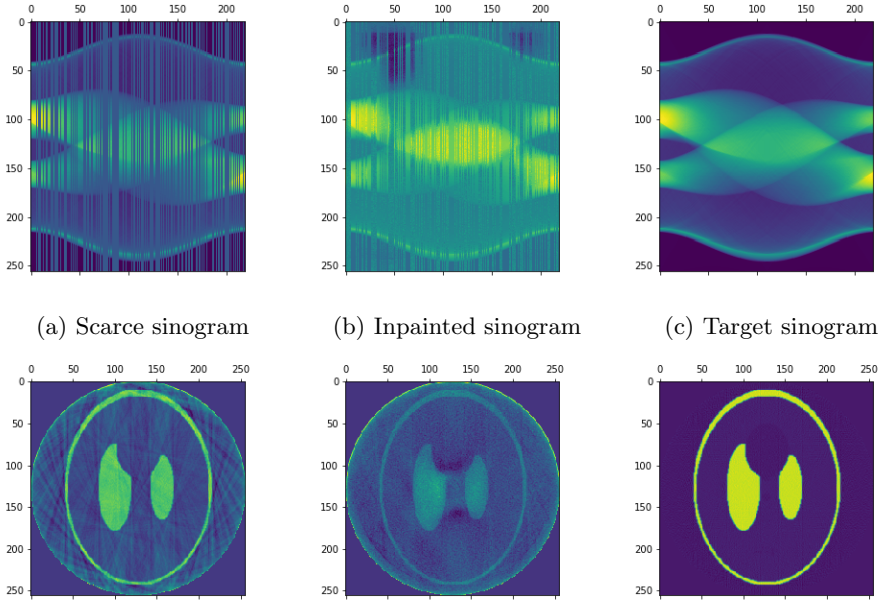
The hierarchical inference proposed by Lee shows its limits when it comes to generate samples from a convolutional DBN. As shown in (5), scaling from a RBM to a DBN does not affect significantly the estimation of missing measurements. We believe that this comes from the fact that the first layer's visible units must be clamped with a value to perform the generation of measurements. Hence, when too many frames are missing from the sinogram, the signal generated from the top-RBM gets crushed by the signal coming from the bottom one : we have found a two orders of magnitude difference. Also, we have noticed that when running the Monte Carlo Markov Chain (MCMC), the value of signal received by the detection layer quickly exploded, even after as few steps as 4. We tried 3 remedies:

- Re-scaling the visible states between 0 and 1 for each steps.
- Replacing the exponential function by a sigmoid function in the probability of activation of hidden units formula.
- Clamping the value of the signal received by the detection layer.

None of them worked enough to allow a satisfying inpainting. The first two remedies allowed the MCMC to run for longer at the expense of a saturation of the signal received, when the clamping of one of the signal only gave unrealistic reconstructions.

We now want to assess the quality of our inpainting. To do so, we compare the sinograms of an acquisition where half of the frames are missing, the inpainted sinogram and the sinogram of the fully sampled acquisition. To have an understanding of the quality of the inpainting, we also display the inverse radon transform of each sinogram.

Even if some frames have been generated, the result is not artefact free. Now, instead of being dominated by streaking artefacts due to a scarce acquisition, is is more subjected to graininess. Also, one must notice that a round halo has appeared in the reconstructions. This is due to the fact that zero attenuation values have been replaced by different values coming either from the removing of the frames or the sampling process from the RBM. In the first case, the sinogram is renormalised and then frames are removed and replaced by zero-valued frames. In the second case the background of the sinogram does not have a zero value but rather a small one. In both cases, non zero-values appear on the top and bottom borders of the sinogram that are transformed into a circle by the radon transform.



(a) Scarce sinogram (b) Inpainted sinogram (c) Target sinogram

(d) Scarce phantom (e) Inpainted phantom (f) Target Phantom

Figure 6: Comparison of the interest sinograms and their Radon transforms

## 4. Related Work and Outlook

The sinogram inpainting problem has been tackled using different approaches. Either as a numerical problem with XCT-specific error functions (Tovey et al, 2019) or as an image inpainting problem addressed by Generative Adversarial Networks (Li et al,2019). No approach to our knowledge includes the CAD data. Even if our investigation has not given the expected improvements so far, it has led us to successfully use RBM for sinogram inpainting. This experience invites us to use DBN as a feature extractor and work on the features of each modality to generate the missing measurements.

## 5. Discussion

RBMs and DBNs are a good way to learn a distribution in an unsupervised fashion. However, they are hard to train and their implementation is complicated, due to the fact they are graph networks.

For our application, the joint layer design is even more tricky to use and has so far not lead to a significant improvement. Indeed, it is very hard to train this layer, even on feature independent vectors. In our experiments, it happened often that each modality representation on the joint layer was encoded on different bits, making it unsuitable for cross-modal recovery. This is why the training is hard to monitor : even if the observables show that training is achieved, the extracted representation does not help in our situation. A remedy for this could be to add a term to the objective function so that it penalises representations that are not suitable for modality recovering or to add another RBM on the top of the joint one.

Furthermore, the convolutional autoencoder is limited to the generation of measure-

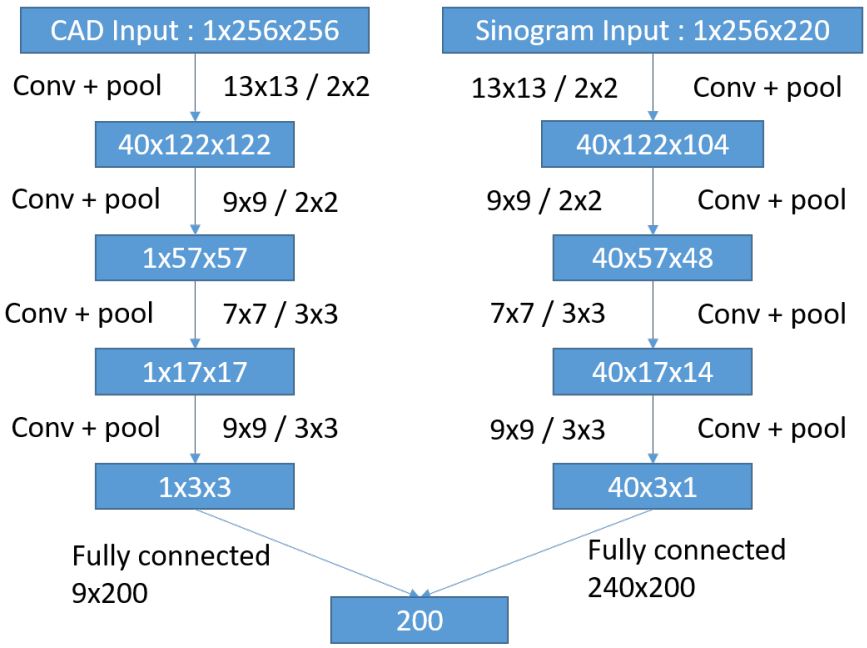



Figure 7: Architecture for feature extraction from two modalities.

ments giving missing measurements, as the unpooling operation in the decoder requires the pooling indexes from the encoder and as the generation of an image from a signal sampled from the top-RBM requires related input on the bottom one.

Supervised fine-tuning tends to produce binary outputs. It can hence be suitable for binary input data but shows limits when it comes to real-valued data as used here. Our understanding of the problem would lead us to think that the shape of the sinogram could be learned from the CAD while its "filling" (the attenuation information it conveys) would come from the scarce measurements. Producing binary sinograms is hence not appropriate to our problem.

5.1. *Acknowledgements*

We would like to thank DSTL and the DGA for supporting this work.

# Appendices

## REFERENCES

Tovey, Robert, Martin Benning, Christoph Brune, Marinus J. Lagerwerf, Sean M. Collins, Rowan K. Leary, Paul A. Midgley, and Carola-Bibiane Schnlieb 2019 Directional sinogram inpainting for limited angle tomography. *Inverse problems* **35.2**, 024004.

Li, Ziheng, Wenkun Zhang, Linyuan Wang, Ailong Cai, Ningning Liang, Bin Yan, and Lei Li 2019 A sinogram inpainting method based on generative adversarial network for limited-angle computed tomography *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine.* **Vol. 11072.** International Society for Optics and Photonics,

Ngiam, Jiquan, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y. Ng. 2011 Multimodal deep learning

Lee, Honglak, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng. 2009 Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations *Proceedings of the 26th annual international conference on machine learning* **pp. 609-616**

Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. 2006 Reducing the dimensionality of data with neural networks. *science 313, no. 5786* **pp. 504-507**

Hinton, Geoffrey E. 2002 Training products of experts by minimizing contrastive divergence. *Neural Computation 14(8)* **pp. 1771-1800**

Jin, Kyong Hwan, Michael T. McCann, Emmanuel Froustey, and Michael Unser. 2016 "Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing 26, no 9* **pp. 4509-4522**

Kelly, Brendan, Thomas P. Matthews, and Mark A. Anastasio. 2017 Deep learning-guided image reconstruction from incomplete data. *arXiv preprint arXiv:1709.00584*

Adler, Jonas, and Ozan ktem. 2017 solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems 33, no. 12* **124007**