



A Reconstruction-Free Projection Selection Procedure for Binary Tomography Using Convolutional Neural Networks

Gergely Pap¹, Gábor Lékó^{2(✉)}, and Tamás Grósz¹

¹ Department of Computer Algorithms and Artificial Intelligence,
University of Szeged, Árpád tér 2, Szeged 6720, Hungary

{papg,groszt}@inf.u-szeged.hu

² Department of Image Processing and Computer Graphics, University of Szeged,
Árpád tér 2, Szeged 6720, Hungary
leko@inf.u-szeged.hu

Abstract. In discrete tomography sometimes it is necessary to reduce the number of projections used for reconstructing the image. Earlier, it was shown that the choice of projection angles can significantly influence the quality of the reconstructions. In this study, we apply convolutional neural networks to select projections in order to reconstruct the original images from their sinograms with the smallest possible error. The training of neural networks is generally a time-consuming process, but after the network has been trained, the prediction for a previously unseen input is fast. We trained convolutional neural networks using sinograms as input and the desired, algorithmically determined k-best projections as labels in a supervised setting. We achieved a significantly faster projection selection and only a slight increase in the Relative Mean Error (RME).

Keywords: Projection selection · Binary tomography · Convolutional Neural Network · Reconstruction-free

1 Introduction

In the field of image processing, the selection of the appropriate projections plays a key role in the reconstruction of binary images. Computed Tomography [4, 8] generates the 2D cross-section images of 3D objects using its projections taken from different directions. The object itself may be regarded as a 3D function representing the X-ray attenuation value at each point of the object, while projections are the line integrals of this function measured on the path of the X-ray beams, turned into a vector. Gathering all the projections from the different directions, we get the sinogram of an object. In most cases, hundreds of projections are needed to produce a high quality reconstruction. However, in some

cases, it is not possible to gather a large number of projections, due to physical and/or time limitations.

Discrete tomography [5,6] employs the assumption that the cross-section image to be reconstructed contains only a few different intensities which are known in advance. This allows us to reconstruct the object with a smaller set of projections and still get an acceptable quality. The purpose of binary tomography is to reconstruct objects containing only one single type of material in a non-destructive way, mostly in industrial cases. However it also could be applied in medical fields for structures such as a particular type of tissue in bones or prostheses. The slices of a binary object can be represented by binary matrices (images), where 1 and 0 denote the presence and the absence of the material, respectively. The range of choice in the case of a small number of projections (say, 20 at most) is quite large.

In [14,19] the authors showed that in most of the cases, the correct selection of the projection angles has a big impact on the reconstruction quality. It means that it is important to find the most informative angles for the reconstruction. There are two main types of projection selection, namely offline and online. In the former case, the sampled angles are known and the projections have already been acquired, i.e. we have a so-called blueprint data [13,18]. In the latter case, the number of projections is not known in advance. However, one can define an upper threshold for the projection number. The adaptive projection selection algorithms allow one to perform dense sampling in the information-rich areas and sparse sampling in the information-poor areas [1–3].

The previously mentioned papers provide a good overview of the available approaches for finding the most informative angles. All of these algorithms focus on solving the problem of projection selection using procedural algorithms. Recently, deep learning approaches, especially Convolutional Neural Networks (CNNs) have achieved tremendous success in various fields such as classification [12], segmentation [15], denoising [20], super resolution [10,16] and removing low-dose related CT noise [9]. Although neural networks are widely used in image processing tasks, in the current literature we could not find any studies that concentrate on how to solve the task of projection selection using machine learning algorithms, or any general process in which CNNs could replace a step regarding reconstruction.

The main aim of this paper is to show that neural networks are capable of solving a complex task like projection selection without any reconstruction step (and to significantly decrease the running time of this process for the online scenario).

The structure of the paper is the following. In Sect. 2, we briefly describe artificial and convolutional neural networks, and in Sect. 3 we outline the methods that we used for projection selection using CNNs. In Sect. 4, we provide details about our experimental frameset. In Sect. 5, we describe how the evaluation of our method was performed, while in Sect. 6 we present the experimental results. Finally, in Sect. 7 we draw some conclusions and make some suggestions for future research.

2 Artificial and Convolutional Neural Networks

Artificial neural networks have provided an efficient and reliable tool for statistical pattern recognition. Neural networks are capable of learning many tasks in the diverse field of computer science and they are also applied frequently to other related disciplines. CNNs generally make use of convolutional layers (2 dimensional in the case of binary images) in which the convolutional neurons respond to a predefined window of perception. Each neuron has its kernel and these kernels are convolved with the image data. After computation with every possible position with its step size, a pooling layer is applied, which computes the maximum (or in some cases the average) of the convolved features. This is necessary to reduce the parameter space and to make the features translation invariant. After the desired number of convolutional and pooling layers, the collected activation values are flattened and connected to a dense neural network, which attempts to solve a classic machine learning task (e.g. classification or regression).

3 Presented Method Using CNN for Projection Selection

Figure 1 summarizes the main steps of our method. The model's architecture consists of 3 convolutional layers with $(5, 5)(3, 3)(3, 3)$ kernel sizes, respectively, followed by two fully connected dense layer with 500 and 180 units. We use ReLU (Rectified Linear Unit) activations as the non-linearities, aside from the last layers' sigmoid function. Between the convolutions we apply maximum pooling with a size of $(3, 3)$. First, the CNN takes a sinogram as input and a dense classifier connected to it outputs 180 activation values optimizing the MSE (Mean Squared Error) using *Adam* (adaptive moment estimation), corresponding to the available directions. Next, these values are thresholded to get the minimum number of projections required for each entity. Since we might end up with more than the required number of projections, a K-means clustering is applied to determine the exact angles to be used in the reconstruction process. We found that the output values of the sigmoid units responsible for angles close to the ground truth are relatively high, which makes clustering them a necessary step. (e.g. often 89-90-91-92 are essential for precise reconstruction, but only one of them should be chosen. Lastly, calculating the RME between the original image and the images reconstructed using the 3 methods explored here (labels, predictions and equiangular projections) gives an estimate of the effectiveness of the selection procedure.

4 Test Frameset

Our image database consisted of 8983 phantoms (icons) of varying structural complexity, each with size 64×64 pixels. To create the train dataset, we performed a modified version of the SFS (Sequential Forward Selection) method [13]. We started the algorithm without initial angles, which resulted in the first

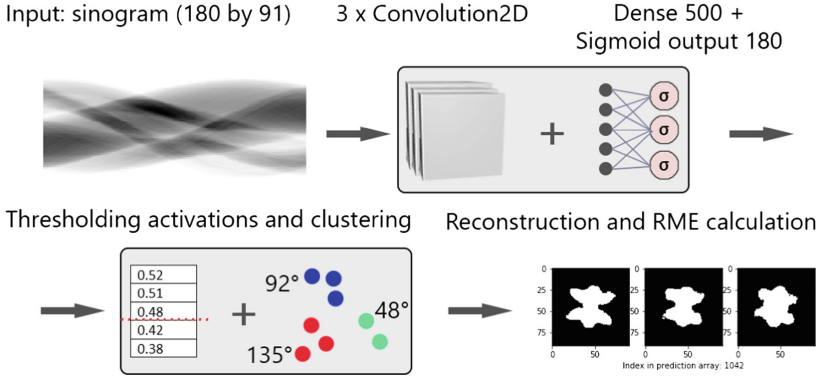


Fig. 1. Our proposed pipeline. The input for the CNN is a sinogram, while the output of the network is 180 activation values (one for each direction) upon which we threshold and use K-means clustering. Afterwards, we reconstruct the images and calculate the RME.

two being selected by the algorithm the same way as it would choose in the case of a bigger number of angles. We did not quite follow the method described in the article, because we ran this algorithm just once, instead of their 18 Multi-start. The reason for these changes is that the SFS's running time was much too long for 8983 images. Furthermore, with this approach we obtained the sequence of the most informative angles. They turned out to be feasible owing to the fact that many of the experiments needed to be done with a different number of label projections (e.g. 4–8). Therefore only having as many projections as needed for the label data could be done without losing the valuable data contained in the algorithmically selected and ordered labels. The labels stored the information in descending order, with the first containing the angle with the most valuable information for minimising the reconstruction error. This way of selecting projections resulted in having only a local optimum with the most informative angles instead of calculating a global one. However, in our case the former was also as applicable to the given task as the latter.

For the projection selection we used the same setup as the authors in [13], except for the above-mentioned changes in the SFS algorithm. For the validation of our CNN, the reconstructions were performed using the thresholded version of the *skimage* python package's SART [4] algorithm. The output of SART is an image with intensity values around 0 and 1. The quality of the reconstructions using the predictions was measured with the Relative Mean Error (RME) defined as

$$RME(\mathbf{x}^*, \mathbf{y}^S) = \frac{\sum_i |x_i^* - y_i^S|}{\sum_i x_i^*}, \quad (1)$$

where \mathbf{x}^* is the blueprint and \mathbf{y}^S is the reconstructed image from angle set S . Our experiments were all performed on 4 NVIDIA Tesla K10 GPUs for equal measurement conditions.

5 Evaluation

Since our input for training consists of sinograms, which can be considered as 2D images, we decided to apply Convolutional Neural Networks. The sinograms extracted from the original images were 91 pixels wide for each projection direction, which formed an 180×91 sized image. The intensity ranges lay between 0–91, so we normalized all of our data by dividing by 91. Reducing input shape by a scaling factor was also experimented with, but we did not notice any improvement regarding accuracy. Based on this observation, we thought about increasing the size of the input parameters, which we also explored using 32760 and 65160 points of data as the source of the training set. These trials were prone to overfitting despite the strict regularisation and normalisation methods applied (dropout [17], batch normalisation [7]). To sum up, 180 by 91 pixels seems to be the optimal size for training the networks, as smaller input features decreased the accuracy of the reconstructions, while larger input spaces led to overfitting (not to mention the increase in training time and memory requirements).

Standard evaluation methods used for scoring neural network predictions might be misleading in the case of a task such as the one presented here. The reason for this is that a label mostly depends on the basic geometrical properties of the input image and it might generate some artifacts resulting from degree-favouritism (selection of common projections such as 0 or 90) or it might find equiangular projections to be the best predictions. The latter might be regarded as the closest one to every possible label projection, but this produces sub-optimal reconstructions. Thus, the images were rotated randomly to eliminate or lessen the effect of the above mentioned problem. Since the main purpose of projection selection procedure is to outperform the Naive equiangular [18] approach by choosing the required angles in order to achieve a lower reconstruction error. To further investigate this issue and to get a better understanding of the selected projections, our method was compared to both the algorithmically selected projection angles, and to the equiangular angle set calculated as $i \frac{180^\circ}{P} \mid i = 0, \dots, P-1$, where P denotes the number of projections used in each case. We will simply refer to the former projection selection method as Label (as it was the training objective of our neural nets) and the latter as Naive, following the authors of [18].

Here, 10-fold cross-validation [11] was used during the training of our network and we based our RME values and other statistical calculations on these runs.

6 Results

In Table 1, we present the RME and Standard Deviation (STD) values of the different methods with 4-6-8 angles, respectively. The \bullet/\circ symbols denote when the differences are statistically significant (\bullet) or not (\circ), using a t-test with a significance level of 0.05. The results are statistically significant only with 8 projections.

The RME values computed from the label projections are naturally the smallest of the three, followed by our CNN approach. The equiangular approach produces the highest RMEs, meaning that the reconstructed images differed from the original ones the most using the Naive angle set. Using our 10-fold cross-validation test evaluations we analyzed the RME values obtained using the 3 methods with 4 projections. We also present our findings in Table 2. In 37 cases our CNN managed to predict a set of projections that was as good as the labels. We note here that these values are from the test set containing examples never encountered during the training of the model. We should add that the more angles we have, the closer we will be to the equiangular approach in terms of minimising the RME. Some of the reconstruction results with the RME values can be seen in Fig. 2 with 8 projections.

Table 1. Average of RME and Standard Deviation values calculated from 10 runs for the three different approaches. The significance values were computed pairwise for Naive-Label, Label-CNN and Naive-CNN and the significant statistical differences are presented column-wise with the symbols of ●/○.

4 projections	Label	CNN	Naive
RME	0.3817 ○	0.4015 ○	0.4912 ○
STD	0.2078	0.2189	0.2903
6 projections	Label	CNN	Naive
RME	0.3196 ○	0.3430 ○	0.3866 ○
STD	0.1723	0.1723	0.2341
8 projections	Label	CNN	Naive
RME	0.2746 ●	0.2940 ●	0.3128 ●
STD	0.1842	0.2174	0.2312

Table 2. The number of images on which the 3 distinct methods gave the smallest RME values are shown along the diagonal. The other cells show where two approaches gave the same RME value.

4 projections	Label	CNN	Naive
Label	4676	37	19
CNN	37	3096	14
Naive	19	14	1141

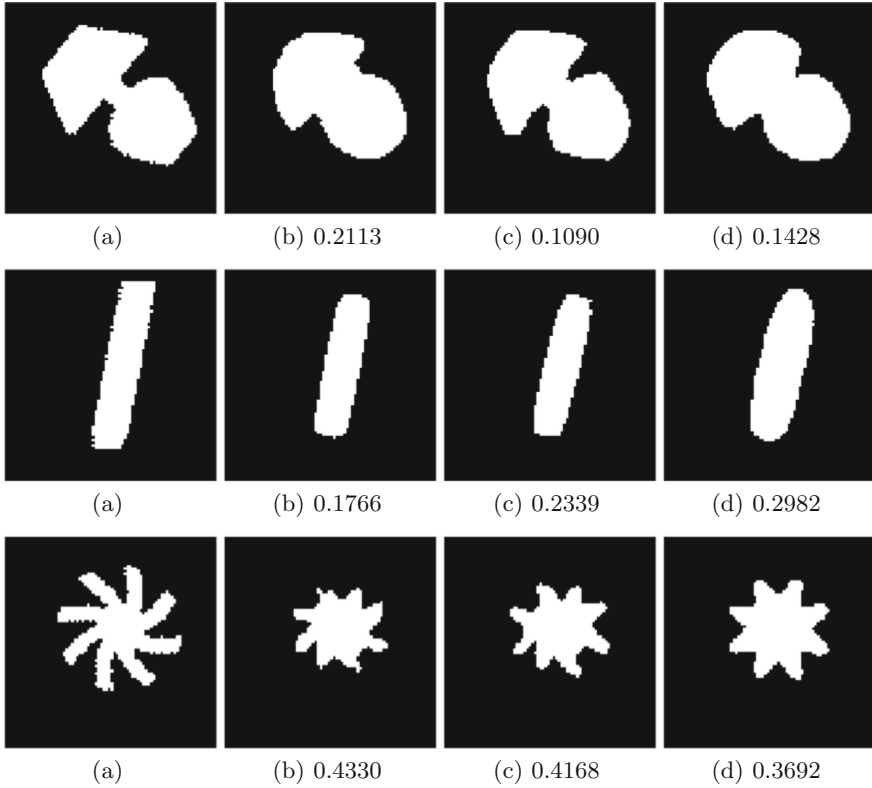


Fig. 2. Reconstruction from 180 projections (a), results of the Label (b), the CNN approach (c) and the Naive approach (d). The values under the figures are the RME values compared to (a). The first row shows a case where the CNN selected the most informative projections. Reconstructing from labels produced the smallest RME in the second row. The last row showcases one image when the equiangular set gave the best score.

7 Conclusions

In this study, we trained a Convolutional Neural Network to select projections for the accurate reconstruction of binary tomography images. After training, prediction is achieved using K-means clustering to get a smaller set of projection data. After comparing the procedural algorithm and the results obtained using a neural network, we observed only a small increase in the RME values compared to the reconstructions from label projections (but performance-wise we achieved a notable improvement with the former). We found that CNNs can be applied to projection selection tasks by training in a multilabel classification scenario. To the best of our knowledge, this was the first attempt to predict projections of binary images for reconstruction using CNNs and to perform projection selection without any reconstruction step. In the future, we intend to

include various projection selection problems and approaches originating from different CT or tomography methods and to evaluate our method using other quality measurements than RME.

Acknowledgements. Gábor Lékó was supported by the UNKP-18-3 New National Excellence Program of the Ministry of Human Capacities. Tamás Grósz was supported by the National Research, Development and Innovation Office of Hungary through the Artificial Intelligence National Excellence Program (grant no.: 2018-1.2.1-NKP-2018-00008). This research was supported by the project “Integrated program for training new generation of scientists in the fields of computer science”, no EFOP-3.6.3-VEKOP-16-2017-0002. The project was supported by the European Union and co-funded by the European Social Fund. We acknowledge the support of the Ministry of Human Capacities, Hungary, grant 20391-3/2018/FEKUSTRAT. The authors would also like to thank István Megyeri for his valuable advice.

References

1. Batenburg, K.J., Palenstijn, W.J., Balázs, P., Sijbers, J.: Dynamic angle selection in binary tomography. *Comput. Vis. Image Underst.* **117**(4), 306–318 (2013)
2. Dabravolski, A., Batenburg, K., Sijbers, J.: Dynamic angle selection in x-ray computed tomography. *Nucl. Instrum. Methods Phys. Res., Sect. B* **324**, 17–24 (2014)
3. Haque, M.A., Ahmad, M.O., Swamy, M.N.S., Hasan, M.K., Lee, S.Y.: Adaptive projection selection for computed tomography. *IEEE Trans. Image Process.* **22**(12), 5085–5095 (2013)
4. Herman, G.T.: *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*. ACVPR, 2nd edn. Springer Publishing Company, London (2009). <https://doi.org/10.1007/978-1-84628-723-7>
5. Herman, G.T., Kuba, A.: *Discrete Tomography: Foundations, Algorithms, and Applications*. Birkhäuser, Basel (1999)
6. Herman, G.T., Kuba, A.: *Advances in Discrete Tomography and Its Applications*. Birkhäuser, Basel (2007)
7. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. *CoRR* abs/1502.03167 (2015)
8. Kak, A.C., Slaney, M.: *Principles of Computerized Tomographic Imaging*. IEEE Press, New York (1988)
9. Kang, E., Min, J., Ye, J.C.: A deep convolutional neural network using directional wavelets for low-dose x-ray CT reconstruction. *Med. Phys.* **44**(10), e360–375 (2017)
10. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016
11. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 14, March 2001
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Neural Information Processing Systems 25*, January 2012
13. Lékó, G., Balázs, P.: Sequential projection selection methods for binary tomography. In: Barneva, R.P., Brimkov, V.E., Kulczycki, P., Tavares, J.M.R.S. (eds.) *CompIMAGE 2018*. LNCS, vol. 10986, pp. 70–81. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20805-9_7

14. Nagy, A., Kuba, A.: Reconstruction of binary matrices from fan-beam projections. *Acta Cybernetica* **17**(2), 359–385 (2005)
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
16. Shi, W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: *Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network*, June 2016
17. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014)
18. Varga, L., Balázs, P., Nagy, A.: Projection selection algorithms for discrete tomography. In: Blanc-Talon, J., Bone, D., Philips, W., Popescu, D., Scheunders, P. (eds.) *ACIVS 2010*. LNCS, vol. 6474, pp. 390–401. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-17688-3_37
19. Varga, L., Balázs, P., Nagy, A.: Direction-dependency of binary tomographic reconstruction algorithms. *Graph. Models* **73**(6), 365–375 (2011). *Computational Modeling in Imaging Sciences*
20. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *Trans. Img. Proc.* **26**(7), 3142–3155 (2017)