Deep neural networks for plasma tomography with applications to JET and COMPASS

Prepared for submission to JINST

3rd European Conference on Plasma Diagnostics
May 6–10, 2019
Lisbon, Portugal

Deep neural networks for plasma tomography with applications to JET and COMPASS

D. D. Carvalho,[1] D. R. Ferreira,[1] P. J. Carvalho,[1] M. Imrisek,[2] J. Mlynar,[2] H. Fernandes[1] and JET Contributors*

EUROfusion Consortium, JET, Culham Science Centre, Abingdon, OX14 3DB, UK
[1]*Instituto de Plasmas e Fusão Nuclear, Instituto Superior Técnico, Universidade de Lisboa 1049-001 Lisboa, Portugal*
[2]*Institute of Plasma Physics AS CR, Prague, Czech Republic*

E-mail: diogo.d.carvalho@tecnico.ulisboa.pt

Abstract: Convolutional neural networks (CNNs) have found applications in many image processing tasks, such as feature extraction, image classification, and object recognition. It has also been shown that the inverse of CNNs, so-called deconvolutional neural networks, can be used for inverse problems such as plasma tomography. In essence, plasma tomography consists in reconstructing the 2D plasma profile on a poloidal cross-section of a fusion device, based on line-integrated measurements from multiple radiation detectors. Since the reconstruction process is computationally intensive, a deconvolutional neural network trained to produce the same results will yield a significant computational speedup, at the expense of a small error which can be assessed using different metrics. In this work, we discuss the design principles behind such networks, including the use of multiple layers, how they can be stacked, and how their dimensions can be tuned according to the number of detectors and the desired tomographic resolution for a given fusion device. We describe the application of such networks at JET and COMPASS, where at JET we use the bolometer system, and at COMPASS we use the soft X-ray diagnostic based on photodiode arrays.

Keywords: Plasma Tomography; GPU Computing; Deep Learning

*See the author list of "Overview of the JET preparation for Deuterium-Tritium Operation" by E. Joffrin et al. to be published in Nuclear Fusion Special issue: overview and summary reports from the 27th Fusion Energy Conference (Ahmedabad, India, 22-27 October 2018).

1 Introduction

Plasma tomography [1] consists in reconstructing the plasma radiation profile on a poloidal cross-section of a fusion device. This reconstruction is based on measurements of line-integrated radiation along multiple lines of sight. There are different types of diagnostics that can be used for this purpose. For example, at JET there is a bolometer system with two cameras (one horizontal and one vertical) providing a total of 56 lines of sight [2], while at COMPASS there are three soft X-ray cameras based on photodiode arrays, which provide a total of 90 lines of sight [3]. The detector technology is different, but in both cases the goal is to reconstruct the plasma radiation profile.

Figure 1 shows the lines of sight for both JET and COMPASS, together with a sample reconstruction. At JET, the horizontal camera has 24 lines of sight, and the vertical camera has another 24, plus 8 extra detectors that can be used as reserve channels. The tomograms at JET have an output resolution of 196×115 pixels. At COMPASS, the three cameras are referred to as A, B and F, with A and B having 35 lines of sight each, and F having 20. This geometry might change from time to time, by adding/removing cameras or changing the lines of sight for existing cameras. The tomograms at COMPASS have an output resolution of 135×105 pixels.

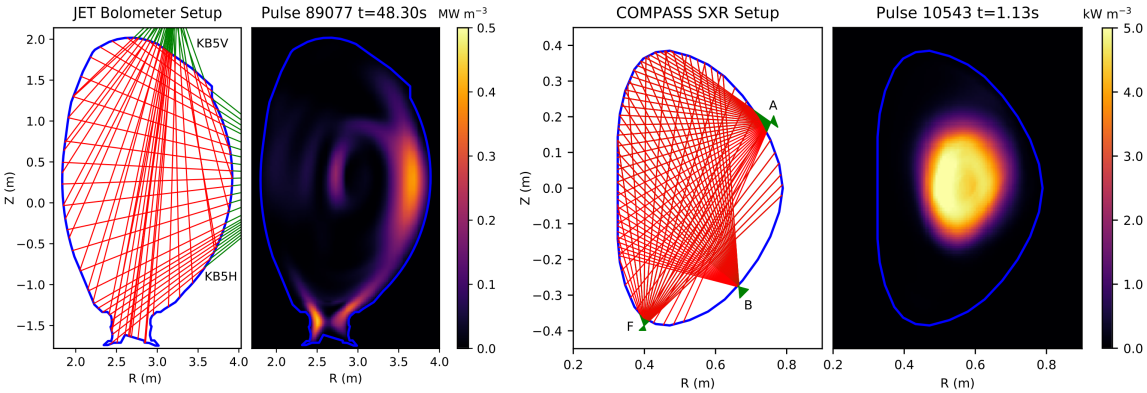

Figure 1. Lines of sight and sample reconstructions for JET (left) and COMPASS (right).

In general, reconstructing the plasma profile from the detector measurements is a computationally intensive task, and several methods exists for that purpose [4]. The method that is used at JET is based on an iterative constrained optimization algorithm that minimizes the error with respect to the observed measurements, while requiring the solution to be non-negative [5]. At COMPASS, the solution is also obtained by minimizing the error to the observed measurements, but with a regularization term that is based on the concept of Fisher information, where the weight of such regularization is found iteratively [3]. In both cases, the methods take a significant amount of time (from a few seconds to several minutes) to converge to a solution.

In this work, we use deep neural networks to produce similar results in a fraction of the time. This approach has already been demonstrated at JET [6], but here we generalize it in order to apply it across multiple devices and diagnostic systems. When running on a Graphics Processing Unit (GPU), such networks are able to produce thousands of reconstructions per second which, considering the sampling rate of these diagnostic systems (on the order of several kHz), is sufficiently fast to provide the prospect of achieving real-time tomography in the near future.

2 Deep neural networks

Deep learning [7] has had a large impact in many areas, especially those involving image processing. In particular, convolutional neural networks (CNNs) have been used for image classification, image segmentation and object detection, to cite only a few examples.

Typically, CNNs comprise a sequence of convolutional layers followed by dense layers. Each convolutional layer applies several filters (in the form of a small kernel or sliding window) to its input. Since the purpose of a filter is to detect a specific feature, its output is called a *feature map*. In addition, CNNs contain subsampling layers. There is usually a subsampling layer after each convolutional layer. (Alternatively, a *stride* greater than 1 may be used in the convolution operator to achieve the same subsampling effect.) The output from these convolutional and subsampling layers is usually a large number of small feature maps. These feature maps are then flattened and connected to dense layers in order to perform image classification based on the extracted features.

In summary, the input to a CNN is typically a 2D image and its output is a 1D vector of class probabilities. However, in plasma tomography, the goal is to produce 2D image of the plasma radiation profile from a 1D vector of detector measurements. For this purpose, we use a network structure that is the logical inverse of a CNN. Such structure has been referred to in the literature as a *deconvolutional network* [8], and it has found applications in image segmentation, object generation and image reconstruction.

Figure 2 shows the general, common structure of the deconvolutional networks that we use for plasma tomography. The network has an input layer with i nodes, which correspond to the number of available detectors (lines of sight). This is followed by two dense layers of size n, which are reshaped into a 3D tensor of size $w \times h \times d$ (so, $n = w \times h \times d$). This 3D shape can be interpreted as comprising d features maps of size $w \times h$. By applying a series of transposed convolutions (with kernel size k), the feature maps are brought up to a size of $8w \times 8h \times d$, from which the output image is generated by one last convolution (with kernel size l). In case the desired output resolution

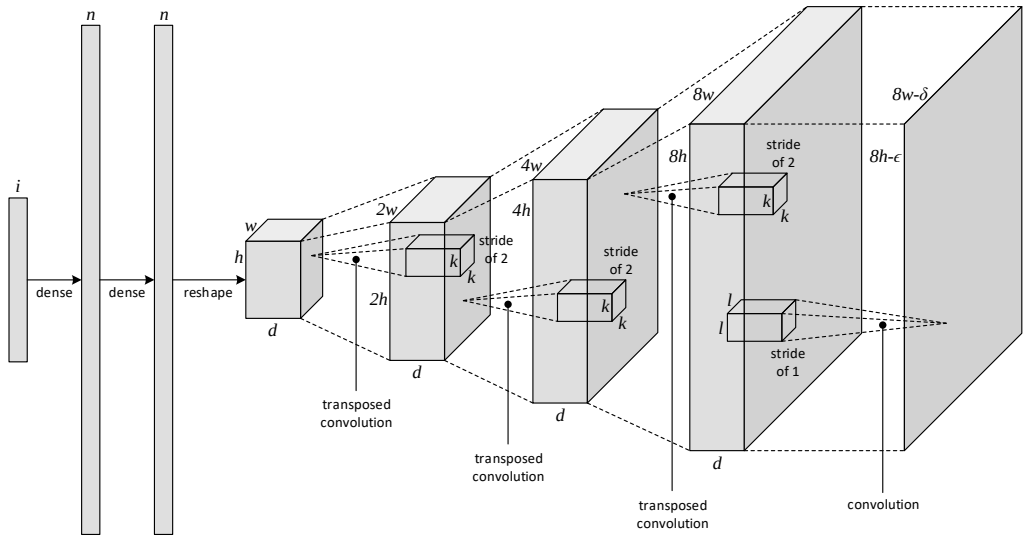

Figure 2. General structure of deconvolutional networks for plasma tomography.

is not exactly $8w \times 8h$, the output shape can be trimmed to $(8w - \delta) \times (8h - \epsilon)$ by a simple slicing operation that is not shown in the figure.

Table 1 specifies the values of the hyperparameters (i, n, w, h, d, k, l, δ, ϵ) for JET and COMPASS. The size n of the dense layers can be set independently of i, but it should be large enough to provide the network with adequate learning capacity for the complexity of the plasma profiles that are to be reconstructed. In this respect, the difference in n for JET and COMPASS does indeed reflect the fact that bolometric profiles at JET are more featureful than soft X-ray profiles at COMPASS. The kernel sizes k and l can be set freely, but typical values are in the range 1–5. Finally, the hyperparameter δ is negative for COMPASS, meaning that, instead of trimming, a zero-padding operation is being applied to increase the width of the output image.

Table 1. Hyperparameters for the trained networks.

	i	n	w	h	d	k	l	$8w$	$8h$	δ	ϵ	$8w-\delta$	$8h-\epsilon$
JET	56	7500	15	25	20	3	3	120	200	5	4	115	196
COMPASS	90	3315	13	17	15	5	1	104	136	-1	1	105	135

3 Datasets and training

To train the network at JET, we gathered about 28000 sample reconstructions from all the campaigns since the installation of the ITER-like wall in 2011. The dataset was divided into 80% for training, 10% for validation, and 10% for testing. Training was done on an NVIDIA Titan X GPU using accelerated gradient descent with a small learning rate (10^{-4}) and a batch size of about 400 samples. For COMPASS, we gathered about 5800 sample reconstructions that belong to an interval of more than 1000 shots performed with the camera geometry in figure 1. The same data splitting and training procedure was followed.

In both cases, the loss function to be minimized was the mean absolute error between network output and the sample tomograms that were provided for training. For JET, the best result was achieved at around epoch 800, with a minimum validation loss of 0.0128 MW m^{-3}. To appreciate the small scale of this error, it can be compared to the dynamic range of the reconstruction in figure 1 (0.5 MW m^{-3}). For COMPASS, the minimum validation loss was 0.0054 kW m^{-3} at around epoch 26000. Again, this value is considerably small when compared to the dynamic range of the reconstruction in figure 1 (5.0 kW m^{-3}).

4 Results

To assess the quality of the reconstructions produced by these networks, we used image comparison metrics such as structural similarity (SSIM) [9] and peak signal-to-noise ratio (PSNR) [10]. Table 2 presents the results achieved on the test sets for JET and COMPASS. In particular, SSIM is close to 1.0 (maximum) and the values of PSNR (in decibels) are roughly equivalent to the error that is introduced when compressing an image to JPEG format. A comparison between original tomograms and those produced by the networks is presented in figure 3.

Table 2. Quality metrics on the test set. Mean and standard deviation values are presented.

	SSIM		PSNR (dB)	
	mean	std.dev.	mean	std.dev.
JET	0.936	0.061	35.36	7.17
COMPASS	0.998	0.004	49.96	4.63

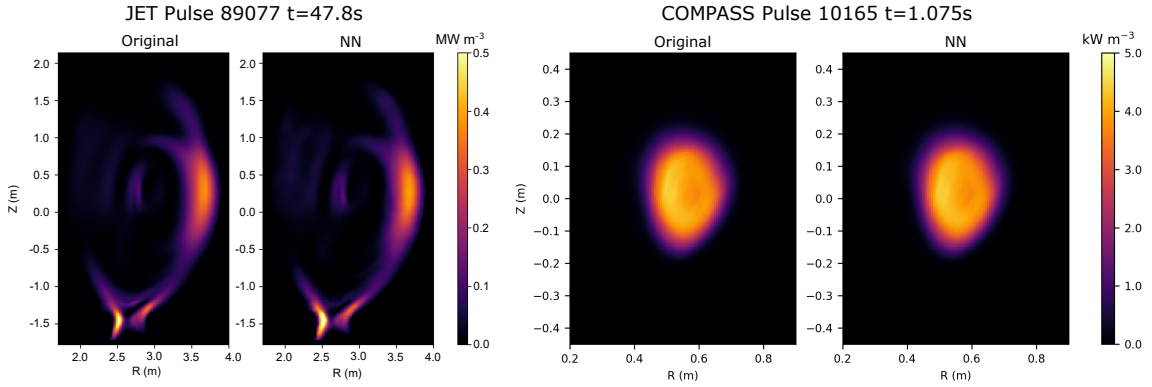

Figure 3. Comparison between original tomogram and network output for JET (left) and COMPASS (right).

In addition to high accuracy, these networks have sufficient performance to produce large batches of reconstructions. When running a GPU, it is possible to compute 3000 reconstructions per second for JET, and 6000 reconstructions per second for COMPASS, since the size of the network for COMPASS (in terms of number of parameters) is basically half of the one for JET. This allows the computation and analysis of full pulse reconstructions, where for example disruptions and their precursors are clearly visible [6].

5 Handling missing detectors

These networks can also be trained to become robust to missing or malfunctioning detectors. For this purpose, we insert a dropout layer [11] with a small dropout rate after the input layer. This way, the network will be forced to learn from a randomly selected subset of detectors, where such random selection is performed continuously during training time.

At test time, we remove up to two detectors by setting them to zero, and again evaluate the network accuracy on the test set. To calculate the evaluation metrics, we average over all possible combinations of missing detectors.

Figure 4 presents the results for four different dropout rates, from zero (no dropout) to 10%. It is possible to observe that, when dropout is not used, the network accuracy degrades with missing detectors. On the other hand, when the network is trained with dropout, it is able to maintain its accuracy when detectors are removed. In addition, a small amount of dropout (5% for JET and 1% for COMPASS) seems to be beneficial even when no detectors are missing.

In conclusion, a small amount of dropout on the input layer does not harm the training process and provides robustness to missing detectors. The dropout rate can be set as an additional hyperparameter of the network that requires proper tunning.

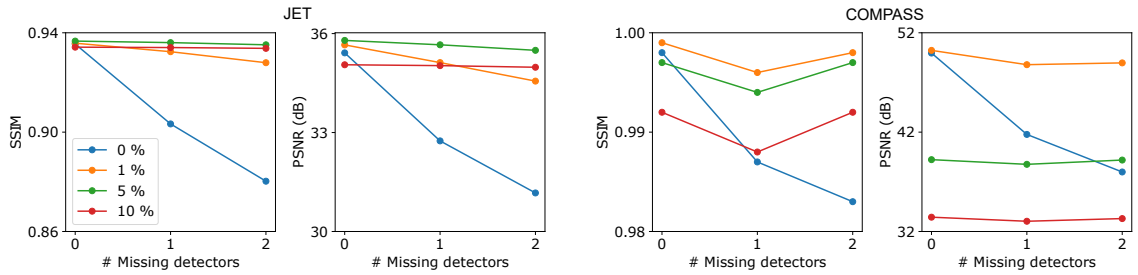

Figure 4. Quality metrics obtained for JET (left) and COMPASS (right) after training with different dropout rates and testing by removing a given number of detectors.

Acknowledgments

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 and 2019-2020 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission. IPFN activities received financial support from *Fundação para a Ciência e Tecnologia* through project UID/FIS/50010/2019. The work developed at COMPASS was co-funded by MEYS project LM2015045 and has received financial support from FuseNet in the scope of its Master student internship programme. The Titan X GPU used in this work was donated by NVIDIA Corporation.

References

[1] L. C. Ingesson, B. Alper, B. J. Peterson and J. C. Vallet, *Chapter 7: Tomography diagnostics: Bolometry and soft-x-ray detection*, *Fusion Science and Technology* **53** (2008) 528–576.

[2] A. Huber, K. McCormick, P. Andrew, P. Beaumont, S. Dalley, J. Fink et al., *Upgraded bolometer system on JET for improved radiation measurements*, *Fusion Engineering and Design* **82** (2007) 1327–1334.

[3] M. Imríšek, J. Mlynář, V. Löffelmann, V. Weinzettl, T. Odstrčil, M. Odstrčil et al., *Optimization of soft X-ray tomography on the COMPASS tokamak*, *Nukleonika* **61** (2016) 403–408.

[4] J. Mlynar, T. Craciunescu, D. R. Ferreira, P. Carvalho, O. Ficker, O. Grover et al., *Current research into applications of tomography for fusion diagnostics*, *Journal of Fusion Energy* (2018) .

[5] L. C. Ingesson, B. Alper, H. Chen, A. Edwards, G. Fehmers, J. Fuchs et al., *Soft X ray tomography during ELMs and impurity injection in JET*, *Nuclear fusion* **38** (1998) 1675.

[6] D. R. Ferreira, P. J. Carvalho and H. Fernandes, *Full-pulse tomographic reconstruction with deep neural networks*, *Fusion Science and Technology* **74** (2018) 47–56.

[7] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning*, *Nature* **521** (2015) 436–444.

[8] H. Noh, S. Hong and B. Han, *Learning deconvolution network for semantic segmentation*, in *IEEE International Conference on Computer Vision*, pp. 1520–1528, 2015.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, *Image quality assessment: from error visibility to structural similarity*, *IEEE Transactions on Image Processing* **13** (2004) 600–612.

[10] Q. Huynh-Thu and M. Ghanbari, *Scope of validity of PSNR in image/video quality assessment*, *Electronics Letters* **44** (2008) 800–801.

[11] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Dropout: A simple way to prevent neural networks from overfitting*, *Journal of Machine Learning Research* **15** (2014) 1929–1958.