UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Corso di Laurea in Fisica

Tesi di Laurea

Evaluation of negative ion beam heat flux profile on STRIKE by neural networks

Relatore

Dr. Gianluigi Serianni

Correlatore

Dr. Rita S. Delogu

Laureando

Anna Steffinlongo

Anno Accademico 2018/2019

Contents

Chapter 1

Introduction

STRIKE (Short Time Retractable Instrumented Kalorimeter Experiment) is an instrumented calorimeter designed to study the SPIDER (Source for Production of Ion of Deuterium Extracted from Radio frequency plasma) negative ion beam. The object of this thesis is the reconstruction of the incident energy flux profile from the thermal image recorded on the rear side of STRIKE. Two different methods based on the use of neural networks have been examined. The goal of the first one is to reconstruct the whole heat flux image, while the second aims to find only the characteristic parameters. The second method has proven to be more efficient and accurate. The comparison between its results and the data obtained by a calorimetric analysis of the problem has shown its reliability.

1.1 Nuclear fusion

In the last decades, the search for clean and sustainable sources of energy has become a fundamental goal. Nuclear energy is at the moment the most promising area; in fact the energy production of several countries is based on nuclear fission. The principle of the production of energy via nuclear fission is that a heavy nucleus splits into more lighter daughter nuclei releasing an amount of energy equal to the difference between the weight of the parent nucleus and the sum of the daughter ones. This energy is released by means of fast neutrons and gamma photons. The most used fuel is ^{235}U which is bombarded by neutrons in order to reach the activation energy. The production of energy via nuclear fission presents some problems such as the scarcity of uranium and the long lifetime of radioactive waste produced by the reaction chains.
The most promising alternative with a high power output is the nuclear fusion where two light nuclei are fused and produce a heavier nucleus provided enough energy is given to overcome the Coulomb barrier [5]. The main reaction studied for the production of energy via nuclear fusion is the fusion of deuterium and Tritium producing a nucleus of helium and a neutron

$$^2D +^3 T \rightarrow^4 He\ (3.5\ \text{MeV}) + n\ (14.1\ \text{MeV})$$

This reaction is particularly interesting due to the low activation energy and the high cross-section, as shown in Fig.1.1. In addition, obtaining the fuel used in this fusion reaction is easier in comparison with uranium. In fact, deuterium is commonly available and Tritium can be obtained from a nuclear reaction involving lithium.

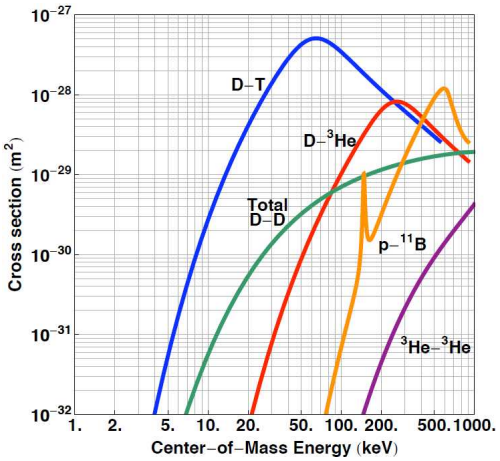

Figure 1.1: Cross-section of several fusion reactions as a function of center of mass energy[1]

At the temperature needed for the fusion reaction to take place the matter is in a plasma state, which means it is a highly ionized quasi-neutral gas of charged particles showing a collective behaviour. In order to produce energy, it is necessary to exceed the break-even point where the energy produced is equal to the energy needed for the reaction to start. Besides, there are continuous energy losses due to transport and radiation phenomena. The Lawson criterion establishes a threshold over which this condition occurs:

$$n\tau_e \geq \frac{12k_BT}{E_\alpha\langle\sigma v\rangle}$$

where n is the plasma density, τ_e is the energy confinement time[2], a parameter that describes the characteristic time of the energy losses, T is the plasma temperature, E_α is the energy of the α particles produced by the reaction and $\langle\sigma v\rangle$ is the reaction rate coefficient. A way to increase the product $n\tau_e$ is increasing the confinement time magnetically confining the plasma. The most used and promising magnetic fusion device is the Tokamak.

[1] https://physics.stackexchange.com

[2] $\tau_e = \frac{W}{P_L}$ where W is the total energy stored in the plasma and P_L is the power lost due to transport phenomena.

1.2 ITER project

In Cadarache (France), a collaboration of 35 nations is building the world's largest Tokamak (ITER) designed to prove the feasibility of using nuclear fusion as an energy source [1]. ITER is expected to produce 500 MW of fusion power for pulses of 400 s using 50 MW for heating the deuterium-Tritium plasma to 150 million °C, obtaining a gain factor of 10. Ohmic heating is not sufficient to bring plasma to such a temperature, so two external heating systems are required. The first one is via RF waves: high-frequency electromagnetic waves propagate in plasma causing the charged particles to oscillate and thus to collide increasing their temperature.
The second consists in neutral beam injection: a beam of high energy neutral particles are injected into the confined plasma. Outside the Tokamak, charged Deuterim particles are accelerated. Subsequently it is necessary to neutralize them so that they can pass through the confinement magnetic field without being deflected. These high-velocity neutral particles are injected in the plasma where they collide with the plasma particles transferring their energy and becoming ionized. ITER will have two neutral beam injectors each one injecting a deuterium beam of 16 MW with particle energies of 1 MeV for pulses of up to 3600 s. For diagnostic purposes, a third neutral beamline will inject a hydrogen beam of 100 keV, 1.5 MW.

1.3 STRIKE

In order to verify the neutral beam injection reliability the PRIMA facility (Padova Research on ITER Megavolt Accelerator) has been built [10]. It contains the two devices MITICA (Megavolt ITER Injector and Concept Advancement) the full-scale prototype of the entire neutral beam injector, still under construction, and SPIDER (Source for Production of Ion of Deuterium Extracted from Radio frequency plasma) that is designed to test and optimize the performance of the ITER-size radio-frequency negative ion source. A SPIDER 3D model is shown in figure 1.2.
SPIDER is equipped with a wide set of diagnostics in order to characterize its ion beam (4 A, 100 keV, formed by 16 groups of 5x16 beamlets), in particular the instrumented calorimeter STRIKE (Short Time Retractable Instrumented Kalorimeter Experiment).
STRIKE is made of 16 one-dimensional Carbon Fiber Composite (CFC) tiles, which are arranged in two panels perpendicular to the beam and intercept the whole beam as shown in figure 1.2. The tiles are not cooled, so STRIKE can be used only for short time operations (< 10 s)[4]. STRIKE was designed to study the beam uniformity and the beamlet divergence, by recording the current passing through the tiles and the temperature image at different distances from the source. Since on the front side the temperature measure would be disturbed by derbis detached from the calorimeter surface and by an optically emitting layer caused by the excitation of the surrounding gas, the temperature is recorded by a set of two infrared cameras (IR) on the rear side of the tiles. This is possible thanks to the fact that the thermal conductivity of the tiles along the beam direction is much higher than that in the perpendicular plane, so the temperature pattern on the rear side is a quite precise reproduction of the one on the front surface. At the moment only half of STRIKE

is in operation. In the analysis reported in this work, the tiles are enumerated as in figure 1.2 and the upper camera is labeled as camera 1.
This thesis proposes a Neural-Network based approach to determine the energy flux profile hitting the front side of STRIKE from the 2D temperature map measured on the rear side in a steady state.

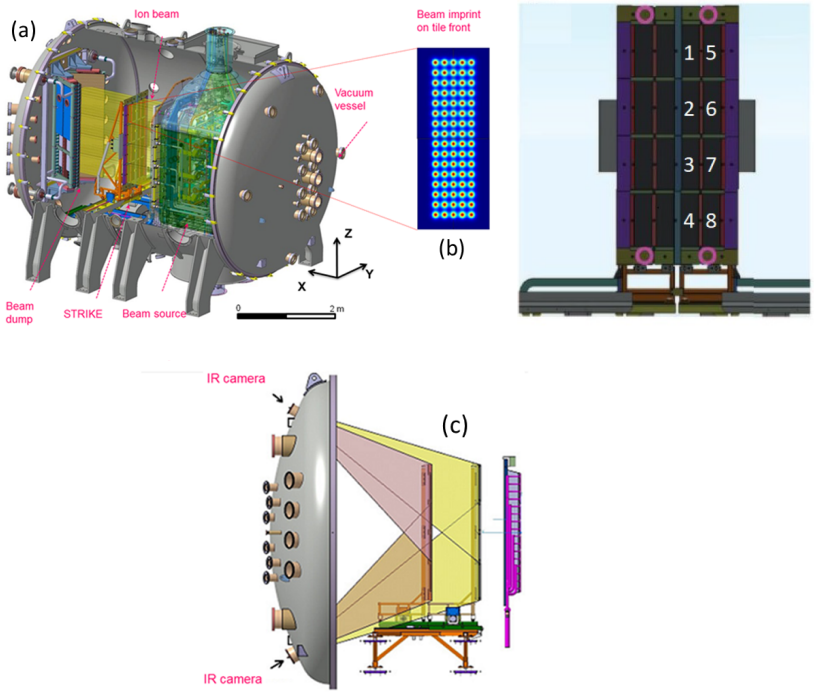

Figure 1.2: STRIKE position inside the SPIDER vessel (left), tiles enumeration [3] (right) and the position of the IRs (bottom)

1.4 Neural networks

A neural network (NN) is a collection of elementary units called neurons [6], [7], [9]. In the last decades, several types of NN have been developed as shown in figure 1.3. To do predictions, the feed-forward NNs are the most used. In feed-forward networks the neurons are arranged into layers: an input layer, some hidden layers and an output layer. In the input layer, the values of the NN input are assigned to the neurons. In a fully connected neural network, each neuron belonging to a layer is connected to every neuron in the previous layer. Each connection between two neurons is associated with two parameters: the weight and the bias. For ease of reference both the weights and the biases will be included in the same parameter vector, w. The neurons in the hidden layers apply an activation function to the weighted sum of the output of every connected neuron from the previous layer:

$$y = f\left(\sum_i w_i x_i\right)$$

The output layer has a number of neurons equal to the number of variables to be determined and the set of their values is the output of the NN. In this work, a Multi-layer perceptron (MLP) with only one hidden layer has been used. As for the activation function, the Elliot sigmoid[3] has been chosen, since this function is fast to be calculated.

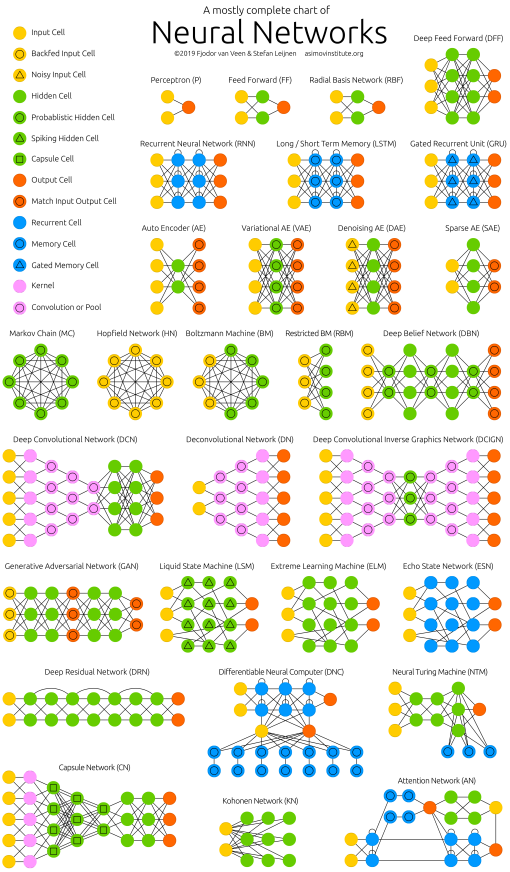

Figure 1.3: Schematic representation of several neural networks[4]

Training

In supervised learning, a set of examples in the form $(input, label)$, where the label is the theoretical output associated with the input, is given to the NN. Training the network means modifying the parameters so that the difference between the label and the network output is minimized. This algorithm is called back-propagation and to do so, it is necessary to define a cost function that represents this difference and to decide an iterative algorithm that allows to optimize the parameters. The easiest back-propagation algorithm is the gradient descent: it allows to find the minimum

[3] $f(x) = \frac{x}{1+|x|}$

[4] Van Veen, F. & Leijnen - asimovinstitute.org

of a given function taking steps proportional to the opposite of the gradient. In the neural network training case it starts from a certain initial set of values w_0 of the parameters, usually randomized, and it takes steps so that

$$w_{n+1} = w_n - \eta_n g_n$$

where $g_n = \nabla_w f(w_n)$ and η_n is the training rate, a parameter that determines the size of the step. Usually η can be a fixed value or a decreasing function of the elapsed time. Using a fixed training rate may be problematic since a too low rate means a really slow convergence and a too high rate may lead to overcome the optimal set of values.
A typical cost function used is the mean square error (MSE)

$$\mathrm{MSE} = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} (\hat{x}_i - x_i)^2$$

where N is the number of examples in the training set, x_i is the network output and $\hat{x_i}$ is the corresponding label. The optimization algorithm used in this work is the Scaled Conjugate Gradient [8], a method similar to the gradient descent, with a converging time by far shorter. The networks used in this thesis have been designed and implemented using the Deep Learning Toolbox for MATLAB®.

Training set

It is really important to wisely decide which examples to include in the training set since they must be representative of examples found in the test set. In this thesis, in order to decide which examples should be in the training set, the geometrical distribution of the samples has been considered. The ones corresponding to the vertices of the linear envelope have been assigned to the training set. All the other elements, that can be expressed as a convex combination of the vertices, have been temporarily assigned to the test set. The determination of the vertices has been performed by means of the Two-Phase Method of Linear Programming [2]. After a complete training of the neural network has been performed using the obtained set, for each element of the temporary test set the mean square error of prediction is determined

$$\mathrm{MSEP} = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \left(\frac{\hat{x_i} - x_i}{\hat{x_i}} \right)^2$$

where N is the number of elements in the actual test set, x_i the network output value and $\hat{x_i}$ the theoretical output. The test element with the greatest MSEP has been moved to the train set, the network has been trained again with the new training set and the MSEP on each test element has been calculated again. These operations have been repeated until the test set reached the 30% of the total number of samples in the database.

Chapter 2

First method: Training the network by using pairs of temperature and energy flux

The first considered approach consists in training the network by using the temperature profiles as input and the associated energy flux profiles as output. Since the heat fluxes associated with the experimental temperature images are not known, it has not been possible to use them to train the neural network, so it has been necessary to build a synthetic database.

2.1 Data set

The database has been created so that it was as similar as possible to the experimental images. The database is made of 2636 pairs of images of 170x70 pixels each, for camera number 1.
The simulated heat flux profiles are caused by five steady-state 3D gaussian beamlets in the power range 5 W − 50 W per beamlet, lasting 2 s, hitting the calorimeter surface. The position of the centers of the beamlets has been chosen according to the position of the temperature peaks on the STRIKE images. In each image, every gaussian has the same Half Width at Half Maximum (HWHM), but in different images the HWHM changes in the range 5 mm − 34 mm. In the by a small angle with respect to the xy axis system are also included. The relative height of the gaussian peaks reproduces the range of variation of the experimental data.
For each heat flux, the corresponding temperature profile has been determined by means of a Finite Element Method (FEM) model implemented in COMSOL®, considering the initial temperature of the calorimeter $T = 300$ K. In the model, the main properties of the tiles have been taken into account, such as the thermal conductivity that in the beam direction is 20 times higher than in the plane normal to the beam direction.
Both for the experimental data and the simulated ones the image recorded at 2000 ms

has been chosen as a reference, because the longer the time, the larger the thermic spots, which become more and more different from the heat pattern, but in the earlier images the temperature difference is too low to be appreciated.

2.2 Pre-processing

Before training the neural network, it is necessary to pre-process the images while retaining only the most meaningful features in order to reduce their volume. Moreover, the pre-processing phase drastically reduces the noise, thus increasing prediction reliability and reducing the risk of overfitting. The processed images are then normalized in the range [-1,1].
Firstly the frequencies are studied by applying a two-dimensional Fast Fourier Transform (FFT) to every image. Thanks to the FFT symmetries, to avoid replicating information some components can be eliminated and from the remaining ones the most relevant frequencies, according to their amplitude, can be selected.
The filtered images can then be processed by means of Principal Component Analysis (PCA) selecting only the most important features.
The optimal number of frequencies and principal components (PCs) is a compromise between the data volume reduction and the precision of the reconstruction. In order to determine it, for different numbers of frequencies and PCs the reconstruction error has been calculated.
The reconstruction error has been defined as the mean square error of prediction (MSEP) by comparing pixel by pixel the original image and the filtered one. The latter has been reconstructed by applying the inverse PCA followed by the inverse two-dimensional FFT.
This procedure has been adopted both for the heat flux and the temperature profiles separately. The images in figure 2.1 represent the MSEP as a function of the number of PCs for a fixed number of frequencies both for the temperature and for the heat flux of the same database sample.

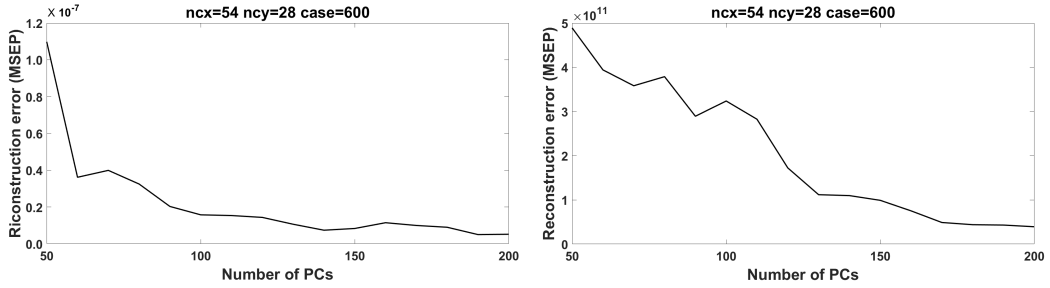

Figure 2.1: MSEP for a temperature (left) and a heat flux (right)

For the sake of simplicity, the same number of principal components has been used for both sets, even though the reconstruction error is considerably lower for the temperature profiles. In particular, 150 principal components have been kept since the greater the number of components, the greater the noise passed on to the neural network. The graphs in figure 2.2 report the comparison between the original and

the filtered profiles, when plotting the section profile along the line passing through the maximum of the original image.
The obtained filtered data will be normalized in the range [-1,1] and divided into training and test sets as explained in chapter 1.4 on page 6, by calculating the MSEP between the reconstructed and the target images.

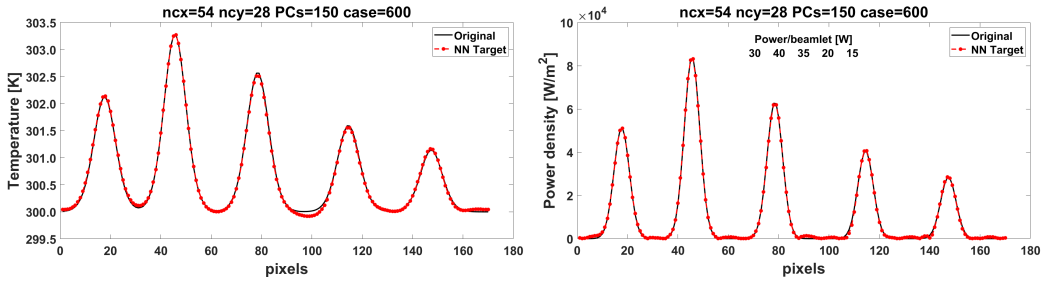

Figure 2.2: Comparison between the original and the filtered temperature profile (left) and heat flux profile (right)

2.3 Training

A set of about five neural networks has been trained. The first NN has been built with a number of neurons in the hidden layer equal to the number of PCs, while the others have been built by adding in the hidden layer, 10 neurons each time.
The best one has been determined as follows. For each network, the mean square error (MSE) on the reconstructed images of the definitive test set have been calculated pixel by pixel. The greatest MSE of each network has been compared and the NN with the lowest one has been declared the best NN. The target, the reconstructed flux profiles and the section profile long the line passing through the maximum of the original one are reported for the best and the worst test cases. In the best case the network is able to reconstruct the centroid positions quite accurately, but the amplitude and the Half Width at Half Maximum are not correct. On the other hand, in the worst case the reconstruction presents deformed peaks with none of the characteristics of the target image correctly reproduced. Since for every test case the reconstructed image is fairly different from the target one, it is possible to conclude that the expected results are not properly achieved by the network.

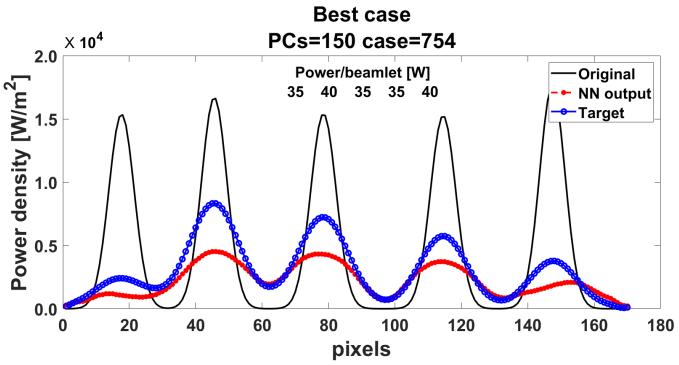

Figure 2.3: Best-case comparison between the original image, the target one and the one reconstructed by the neural network (NN)

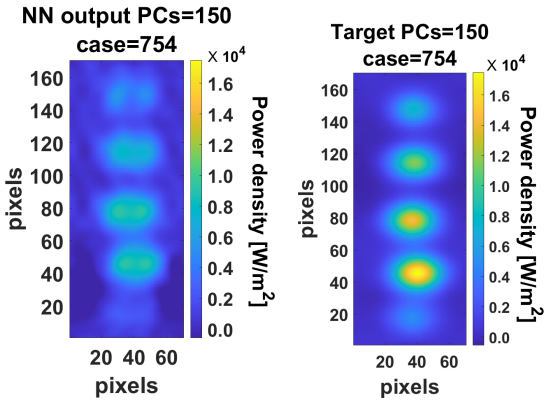

Figure 2.4: Flux profile reconstructed by the NN (left) and target profile (right) for the best case

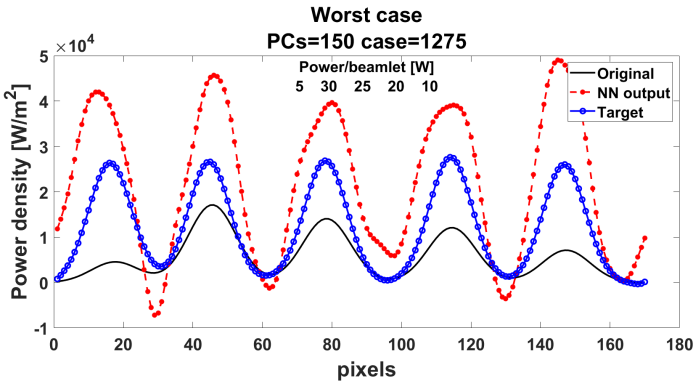

Figure 2.5: Worst-case comparison between the original image, the target one and the one reconstructed by the network

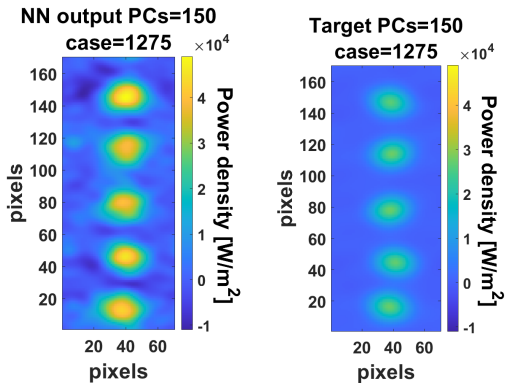

Figure 2.6: Flux profile reconstructed by the NN (left) and target profile (right) for the worst case

2.4 Experimental data

The data here analyzed are the shots in the range [6229, 6275] recorded during the 2019 experimental campaign. Since the cameras do not face frontally the calorimeter, the images have been preliminarily elaborated to correct the perspective.
In order to reduce the noise at the border of the images, it is required to apply a filter to the experimental images. After trying with several filters, such as the smoothing and the radial basis function fit, the gaussian fit has proved to be the best one. So firstly the images have been filtered by a gaussian fit, then they have been pre-processed in the same way as the temperature profiles of the database by selecting the most important frequencies and projecting the images onto the space generated by the temperature principal components. Finally, they have been normalized and given as an input to the best network.
Two examples of the flux reconstructed by the neural network are reported in the following images.
With a computer having a Intel® Core™ i7-8565U CPU and 16 GB of RAM, training a network with this method takes approximately 5 hours. The results obtained are not acceptable, so a second approach has been examined.

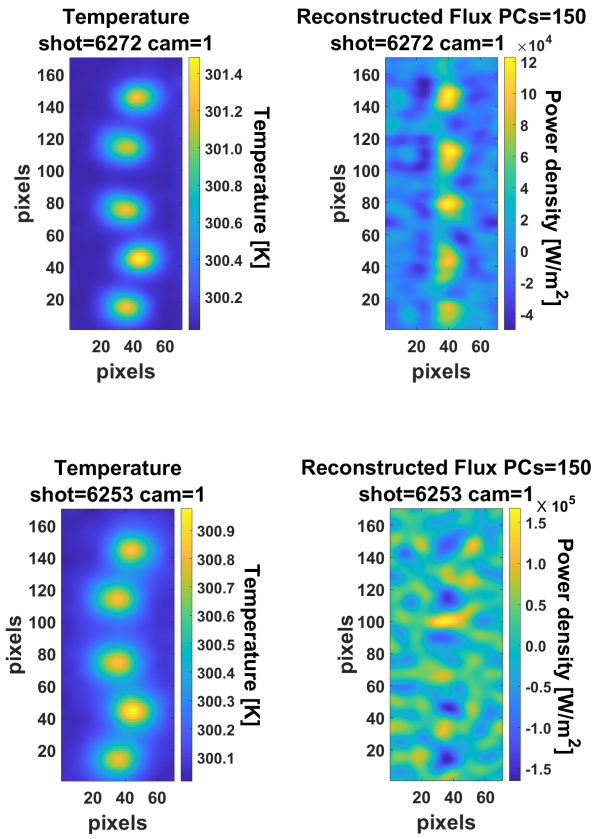

Figure 2.7: Temperature profile post gaussian filter (left), reconstructed heat flux for two different shots (right)

Chapter 3

Second method: Training the network by using pairs of temperature and descriptive parameters of the energy flux

The second approach consists in training the NN using the same input set as before, but changing the output one. Instead of heat flux images, the output set is formed by vectors containing the parameters of the heat flux gaussians. The parameters considered are the peak amplitudes, the centroid positions and the Half Width at Half Maximum along the two axes.

3.1 Pre-processing and training

As for the pre-processing, the temperature images have been treated as before, while the parameter vectors needed only to be normalized. Since, even using only a limited number of principal components, the reconstruction error on the temperature images is not so large, only 40 PCs have been taken into account. The figure 3.1 reports the temperature profile section through the maximum for the same case as in figure 2.1 with 40 principal components.

As in section 2.3, a set of neural networks has been trained and with the same criterion, the best network has been decided. It is interesting to observe that since the number of components is 40 instead of 150 the networks are significantly smaller and therefore the training time is rather shorter. Indeed the same computer used before took approximately one hour per network to train.
The best and the worst cases are reported in figures 3.2, 3.3 and 3.4, 3.5. While the best case is perfectly reconstructed, the worst one presents a light underestimation of the peaks amplitude. As far as the test cases are concerned, the network exhibits a good reconstruction capability.

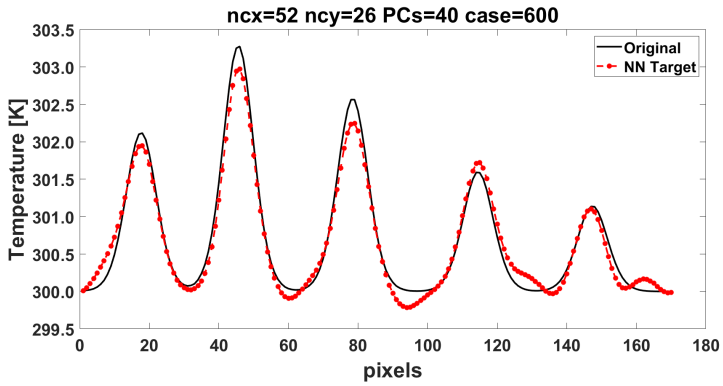

Figure 3.1: Comparison between the original and the filtered temperature with 40 PCs

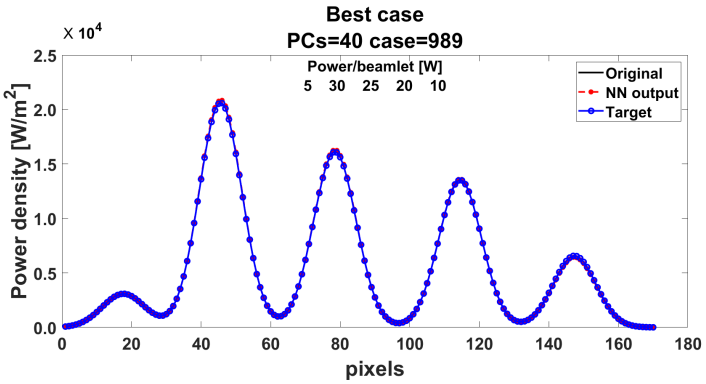

Figure 3.2: Best-case comparison between the original image, the target one and the one reconstructed by the neural network (NN)

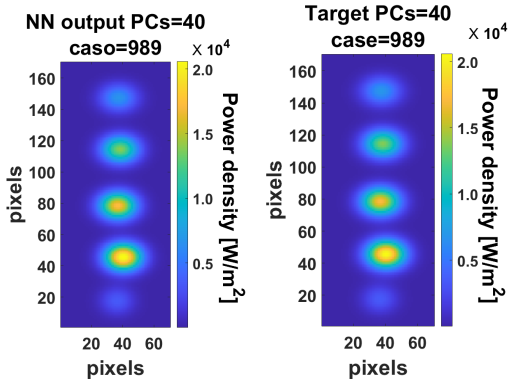

Figure 3.3: Flux profile reconstructed by the NN (left) and target (right) for the best case

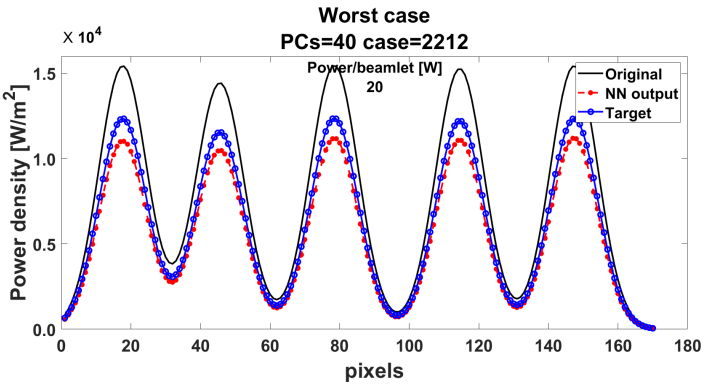

Figure 3.4: Worst-case comparison between the original image, the target one and the one reconstructed by the network

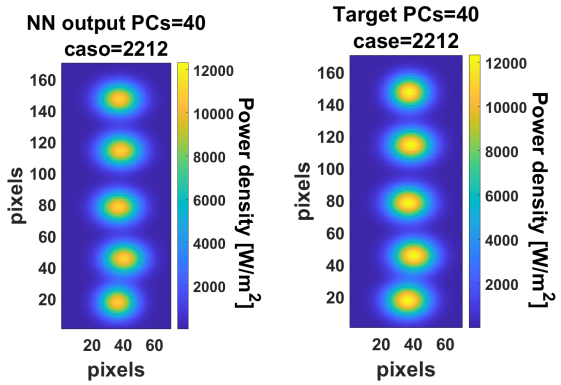

Figure 3.5: Flux profile reconstructed by the NN (left) and target (right) for the worst case

3.2 Experimental data

Regarding the experimental data, the same procedure as before has been followed. The same cases are now reported.

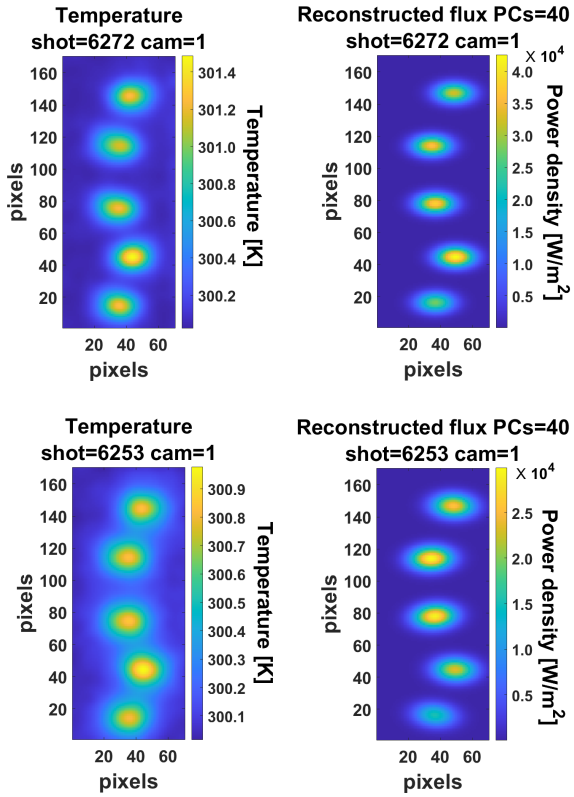

Figure 3.6: Temperature profile post gaussian filter (left) and the reconstructed heat flux (right) for the same examples as before

Since the output of the network is associated with the heat flux, in order to verify the reconstruction accuracy, it has been necessary to simulate the temperature profile caused by it and compare the latter with the experimental data. This simulation has been performed by the COMSOL® based Finite Element Method model used to prepare the database. The simulated temperature profile has been compared with the original one and the one obtained by applying the gaussian filter as reported in the figure 3.7.
Moreover, the total power has been estimated from the reconstructed heat flux and it has been compared to the power obtained from a calorimetric analysis. In figure 3.8, the comparison between the two estimes of the total power for each experimental sample is reported. From this comparison, it is remarkable that the NN tends to underestimate the power fluxes.

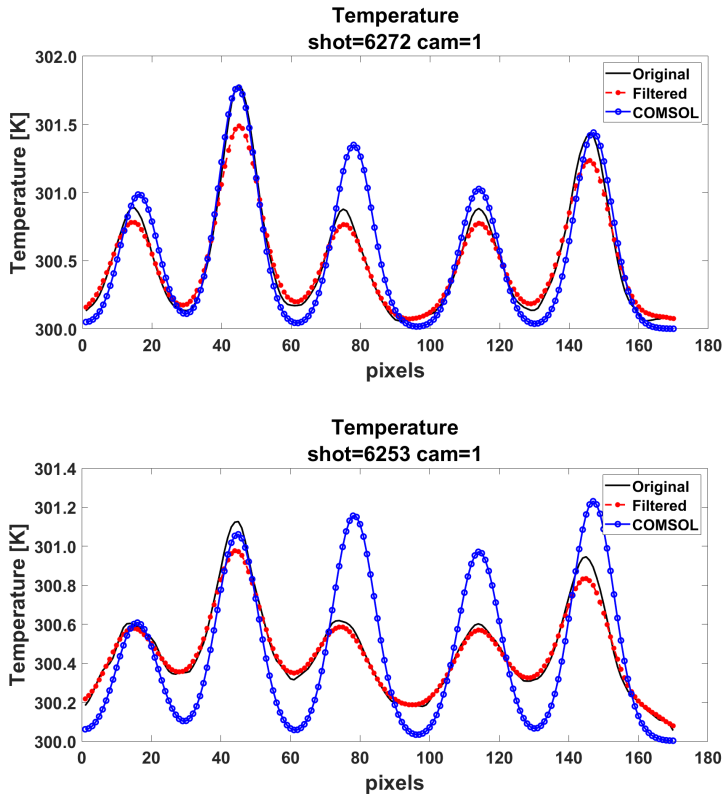

Figure 3.7: Comparison between the original temperature, the gaussian filtered one and the one obtained from the COMSOL simulation for two example shots

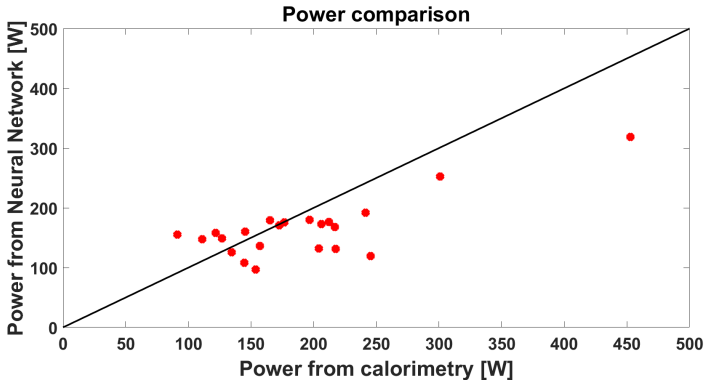

Figure 3.8: Comparison between total power obtained from the neural network heat flux reconstruction and the calorimetric analysis for each shot

To furtherly test the precision of the reconstruction and investigate the power underestimation, the temperature images simulated via COMSOL® were filtred using a gaussian fit and the parameters of the fit have been confronted with the ones associated with the experimental data. An example of this comparison is shown in figure 3.9, where on the x-axis the parameter values from the reconstruction are reported, while the ones obtained from the experimental data are reported on the y-axis. The more the two sets of parameters are consistent, the more the points will

be distributed on the bisector of the quarter. From this comparison it is possible to observe that the NN gives good results as far as amplitude and centroid positions are concerned, but is not able to find the right half width at half maximum (HWHM) in either of the two axis. This is due to the fact that the HWHM is fixed for each example used to train the network, thus the NN is not capable of recognizing the variation of this parameter inside the same image. This network incapability could play a role in the power underestimation.

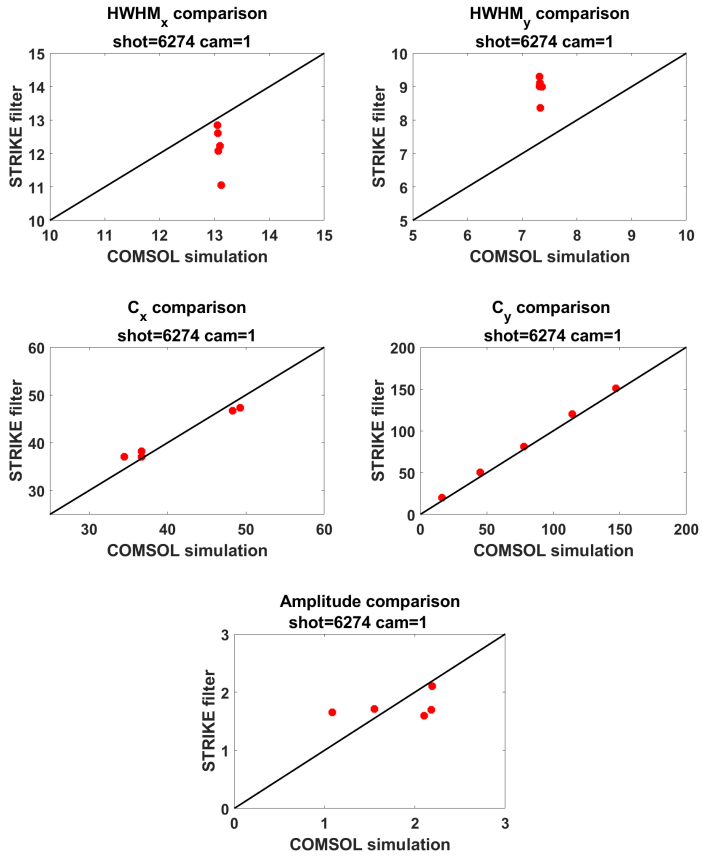

Figure 3.9: Comparison between the gaussian parameters of the filtered image and the one simulated by COMSOL

Chapter 4

Conclusions

STRIKE is the diagnostic system devoted to the spatially resolved characterization of the negative hydrogen beam produced by the accelerator of SPIDER, the prototype negative ion source for heating and current drive in ITER.
The ultimate goal is the implementation of the final method to analyze the thermal images measured with STRIKE by reconstructing the features of the energy flux associated with the SPIDER beamlets in real-time.
In this work, in order to reconstruct the heat flux profile hitting the first tile of STRIKE from the temperature map measured on the rear side, two different neural network based approaches have been examined. The first considered approach consists in training the network with pairs of heat flux and temperature profiles. This method has proved to be more expensive in terms of computational power and its results are far from acceptable. On the other hand, the second approach, which consists in using only the descriptive parameters of the energy flux instead of the whole profile to train the network, presents really promising results while also exhibiting a lower computational cost. In particular, the latter neural network reconstructs with a good approximation the amplitude and the positions of the gaussian peaks. Nevertheless, the network is not able to find the correct Half Width at Half Maximum (HWHF) for any peak. From the only 23 experimental images available it seems that the NN tends to underestimate the powers.
Future development of this work will involve the inclusion of examples in which each gaussian has a different HWHF in the training database, in order to make it possible for the NN to recognize the variation of this parameter, thus a better reconstruction of the image itself can be expected. Further work shall also involve a thorougher experimentation of the second approach with images in a wider power region in order to tackle the problem of the underestimation.

Bibliography

[1] URL: https://www.iter.org.

[2] Mokhtar S. Bazaraa, John J. Jarvis, and Hanif D. Sherali. "Minimal cost Network Flows". In: *Linear Programming and Network Flows*. John Wiley & Sons, Ltd, 2011. Chap. 9, pp. 453–512. ISBN: 9780471703778. DOI: 10.1002/9780471703778.ch9.

[3] Gloria Canocchi. "First characterization of the SPIDER negative ion beam by the diagnostic calorimeter STRIKE". Università degli Studi di Padova, Tesi di Laurea in Fisica. 2019.

[4] Rita S. Delogu et al. "Neural network based prediction of heat flux profiles on STRIKE". In: *Fusion Engineering and Design* 146 (2019). SI:SOFT-30, pp. 2307–2313. DOI: https://doi.org/10.1016/j.fusengdes.2019.03.178.

[5] *Fusion Physics*. Non-serial Publications. Vienna: International Atomic Energy Agency, 2012. ISBN: 978-92-0-130410-0. URL: https://www.iaea.org/publications/8879/fusion-physics.

[6] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[7] David J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003. ISBN: 9780521642989.

[8] Martin Fodslette Møller. "A scaled conjugate gradient algorithm for fast supervised learning". In: *Neural Networks* 6.4 (1993), pp. 525–533. ISSN: 0893-6080. DOI: https://doi.org/10.1016/S0893-6080(05)80056-5.

[9] J. Schmidhuber. "Deep learning in neural networks: An overview". In: *Neural Networks* (2015). DOI: 10.1016/j.neunet.2014.09.003.

[10] V Toigo et al. "The PRIMA Test Facility: SPIDER and MITICA test-beds for ITER neutral beam injectors". In: *New Journal of Physics* 19.8 (2017). DOI: 10.1088/1367-2630/aa78e8.