

Hyperparameter Optimisation of Artificial Intelligence for Digital REStoration of Cultural Heritages (AIRES-CH) Models

Alessandro Bombini[1(✉)], Lucio Anderlini[1], Luca dell'Agnello[2], Francesco Giacomini[2], Chiara Ruberto[1], and Francesco Taccetti[1]

[1] INFN, Florence Section, Via Bruno Rossi 1, 50019 Sesto Fiorentino, FI, Italy
bombini@fi.infn.it

[2] INFN CNAF, Viale Carlo Berti Pichat 6, 40127 Bologna, BO, Italy

Abstract. Artificial Intelligence for digital REStoration of Cultural Heritage (AIRES-CH) aims at building a web-based app for the digital restoration of pictorial artworks through Computer Vision technologies applied to physical imaging raw data.

In previous work [7], it was shown that it is possible to develop a multidimensional deep neural network capable of inferring the RGB image from an X-Ray Fluorescence raw data.

The developed network comprises two branches: a one-dimensional branch, which works pixel-by-pixel, and a two-dimensional branch, capable of performing image segmentation.

In this project, we report the results of the hyperparameter optimisation of both branches.

Keywords: Machine learning · Cultural Heritage · Nuclear physics

1 Introduction

In recent times, the digital technologies for image manipulation are becoming increasingly relevant in the context of physical technologies applied to pictorial artworks imaging. An interesting example is the application of modern artificial intelligence-based image inpainting to physical imaging [14,16,18,32,36][1].

In a previous work [7], we focused on the application of a multidimensional, multi-spectral deep neural network (DNN) to raw data of physical imaging techniques, as a means to digitally reconstruct non-visible pictorial layer; this was

[1] For other Machine learning approaches in Cultural Heritage, see [9], and references therein.

This research is part of the project AIRES-CH - Artificial Intelligence for digital REStoration of Cultural Heritage (CUP I95F21001120008) jointly funded by Tuscany Region (Progetto Giovani Sì) and INFN.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
O. Gervasi et al. (Eds.): ICCSA 2022 Workshops, LNCS 13377, pp. 91–106, 2022.
https://doi.org/10.1007/978-3-031-10536-4_7

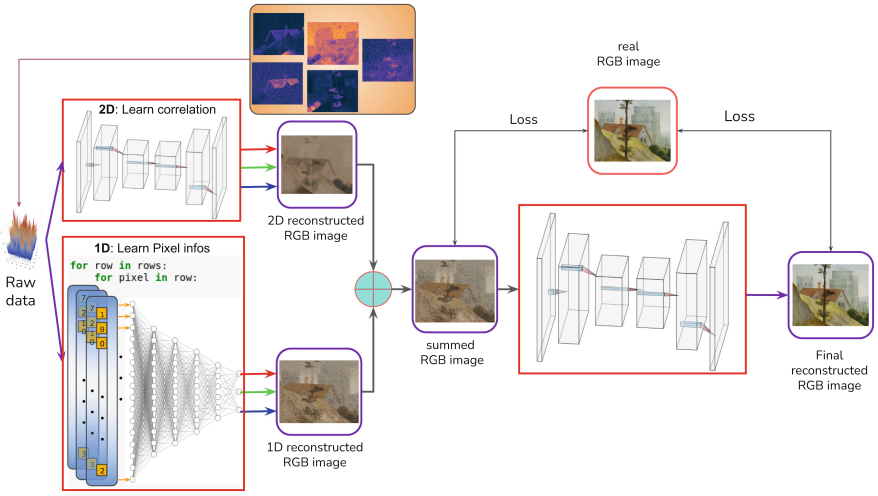

Fig. 1. A sketch of the multidimensional DNN.

done in order to furnish a visual help to restoration professionals through digital technologies.

In [7], we have shown that both branches of the multidimensional DNN we developed were actually capable of inferring an RGB image from a X-Ray Fluorescence raw data. The devised network has two branches: a one-dimensional-branch, trained to infer the colour of each pixel taking the energy spectrum as an input; and a two-dimensional-branch, trained to capture spatial correlations between adjacent pixels.

The results obtained with a custom DNN and several well-known architectures for either the 1D and 2D branches have been compared. The alpha version of the network was also offered in the INFN-CHNet cloud service [8].

In this work, instead, motivated by those results, we employ a different approach: we use *Optuna* [2,3] to actively generate deep neural networks, architecturally similar to the ones developed in [7], and optimise their hyperparameters.

The approach is to use the Optuna capacities of exploring categorical spaces, to let it explore neural networks with different numbers *and* types of layers: either pure Convolutional, Inception [31] and/or (Dilated) Residual blocks [12,35].

After that, we briefly analyse the best model found using *WeightWatcher*, an open-source Python tool for computing quality metrics of trained Deep Neural Networks, based on Random Matrix Theory [20,22,23]; we will show that the two-dimensional branches' layers are either rightly or under-fitted, as expected, leaving room for improvement by enlarging the training dataset. Thus, to increase the size of the dataset, a campaign of new XRF imaging measurements on pictorial artworks of the *Biblioteca Marucelliana* in Florence is scheduled to begin soon, in the context of the joint project AIRES-CH.

Nevertheless, this is just a small step forward. The ultimate goal of the AIRES-CH project, in fact, is to develop a DNN capable of inferring the RGB

image from an XRF image. This is per se useful, since it may give already few information to cultural heritage scientists, but the final goal is broader. Such networks, while providing immediate visual support to physicists performing XRF analyses, represent a necessary building block towards the implementation of algorithms designed to digitally restore hidden (or damaged) pictorial layers, by removing from the XRF scan the components describing the most superficial, visible, layer, in order to furnish a visual aid to CH experts, before the application of any invasive restoration process.

2 The AIRES-CH Project so Far

Artificial Intelligence for digital REStoration of Cultural Heritage (AIRES-CH) is an INFN project devoted to the building of a web application for the digital restoration of pictorial artworks through Computer Vision technologies applied to raw data of physical techniques, such as the X-Ray Fluorescence (the one employed in [7], and in this work).

X-Ray Fluorescence (XRF) spectrometry is a technique suitable for a wide range of applications in many disciplinary fields, providing a fast, sensitive, multi-elemental analysis. It is widely employed in Cultural Heritage (CH) applications, since it allows for non-invasive, non-destructive measurements and no sample pre-treatment is required (see [15,19], and references therein). It is possible to perform a macro mapping, known as Macro X-Ray Fluorescence mapping (MA-XRF), to produce elemental distribution maps of an artwork, proving to be extremely useful for material characterisation, and for understanding the artist's painting and production techniques [5].

Furthermore, and crucially, XRF is able to detect signal coming from hidden pictorial layers; therefore, XRF raw data are suitable to devise a neural network capable of digitally restoring pictorial artworks.

The first step toward such goal, is to develop a neural capable to infer the RGB image from the XRF raw data, seen as a rank-3 tensor $\mathcal{I}_{ij;k}$; the $(i, j) \in [0, \text{width}] \times [0, \text{height}]$ indices refers to pixel's location, while the k index refers to the energy channel in the $(0.5, 38.5)$ KeV portion of the X-Ray spectra (the one explored by our detector)[2].

In order to do so, and prove that is feasible, we developed a *multidimensional* neural network, comprising two branches (Fig. 1):

1. The 1D branch, whose goal is to learn how to associate a 3-vector living in $[0, 1]^3 \subseteq \mathbb{R}^3$, (i.e. the RGB colour) to a vector living in $\mathbb{N}^{500}$ (one for each channels of the detector), the histogram of XRF counts in the energy range;
2. The 2D branch, whose goal is to learn how to associate to 500 grey-scale images a single RGB image.

[2] The tensor $\mathcal{I}_{i,j}$ contains, for each pixel (i, j), the histogram of the counts in bins of the X photon energies. It is customary to convert the raw data into elemental maps by integrating the per-pixel energy spectra around the characteristic energy peak associated to each element [15,19].

2.1 The Training Dataset

The training dataset is the same employed in [7]. It is composed by 62 XRF raw data coming from several XRF analysis on multiple paintings performed both in the LABEC facility in Florence, as well as in situ analysis (the data comes not only from published works, and include some private artworks) [1,4,6,10,24,26,28,29]; for the training of the one-dimensional models, only a 50% of the pixels where employed and randomly chosen, giving a training dataset of around 2 059 780 [histogram, RGB] pairs, divided into training, test, and validation set. Instead, for the training the 2D models, 45 XRF scans are used, reserving the remaining as 9 for test, and 8 for validation.

The raw data comes from different artwork techniques: multi-layered paintings, drawings without preparatory layers, and illuminated manuscripts, all spanning different periods and epochs, from middle ages to contemporary art.

2.2 The Neural Network Models

We have devised several models for each branch. Following a customary choice in Computer Vision, all the networks inferring an RGB colour code are trained using the *Binary Cross Entropy* (BCE) as loss function, while their performance is assessed through several common Image Quality measures [37]: *Peak Signal-to-Noise Ratio* (PSNR), the *Structural Similarity Index Measure* (SSIM) [34] and its variant, the Multi-Scale SSIM (MS-SSIM) [33].

For the 1D branch, we have seen that a custom multi-input DilResNet network slightly outperformed a FractalNet and a WaveNet model; in particular, we noticed that the custom multi-input model performs slightly better in all visual scores, and has the lowest (averaged) Mean Squared Error on the validation images. Instead, the WaveNet mode, which is also the model who has the lowest Binary Cross Entropy error, i.e. the loss function, is the one whose training was longer and more stable. This seems to suggest that future training should be performed using different losses (or metrics).

The peculiarity of the multi-input network relies in the fact its non-topologically trivial, since another input is fed to the Neural Network. In fact, we additionally pass a set of four element peak histograms to a branch of the network: Tin (Sn) Kα, Potassium (K) Kα, Manganese (Mn) Kα, and Titanium (Ti) Kα. Those elements were selected because are usually relevant in pigments, but appears with very small abundance, and thus fewer counts. So we force the network to look carefully for them, by inserting them into a preliminary network.

2.3 Summary of Results

For the 2D branch, we have tested deeply two model architectures of UNet-like shapes [27]: a VGG-based [30] Encoder-Decoder architecture, and a dilated residual UNet network, loosely based on the *Globally and Locally Consistent Image Completion* (GLCIC) network [13]. The summary of the results found are reported in Tables 1a, b, 2b.

Table 1. Performances of the 1D models in [7]. v5 is the custom multi-input model described in the text, v6 is a FractalNet model [17], while v7 is a WaveNet model [25].

(a) Comparison table of the visual scores performances of the v5, v6, v7 1D models.

	SSIM	MS-SSIM	PSNR
v5	0.38844	0.68032	20.10390
v6	0.37207	0.67664	20.07558
v7	0.35640	0.67343	19.97972

(b) Comparison table of the losses performances of the v5, v6, v7 1D models.

	BCE	MSE
v5	0.63645	0.01280
v6	0.63306	0.01407
v7	0.62872	0.01456

Table 2. Performances of the 2D models in [7].

(a) Comparison table of the visual scores performances of the VGG, DilResNet 2D models.

	SSIM	MS-SSIM	PSNR
VGG	0.73290	0.62626	19.69109
DRN	0.74470	0.66946	21.09683

(b) Comparison table of the losses performances of the VGG, DilResNet 2D models.

	BCE	MSE
VGG	0.48037	0.01589
DRN	0.46043	0.01212

The scores of the 2D models appear to be significantly higher w.r.t. the 1D models' scores, as seen in Table 3, but the visual result is less satisfying; as an example, we refer to Fig. 2, where the two best performing models, one for each branch, i.e. the custom multi-input model and the two-dimensional DilResNet model, were applied on one of the validation set raw data, which is a detail of the painting *Putto*, Unknown author, 18th-century (circa) [1].

Table 3. Comparison table of the visual scores performances of the best 1D and 2D models on the detail of *Putto*, Unknown author, 18th-century (circa).

	SSIM	MS-SSIM	PSNR
1D v5_custom multi-input	0.497	0.840	22.048
2D DilResNet	0.818	0.810	25.529

3 Hyperparameter Optimisation

As said in Sect. 1, in this work we follow a different approach to handle the task at hand: starting from the same dataset employed in [7], we decided to employ the open-source python package *Optuna* to fine-tune the DNN hyperparameters;



(a) Integrated XRF Image (b) True RGB Image (c) Recoloured by custom multi-input (1D branch) (d) Recoloured by DilResNet (2D branch)

Fig. 2. Comparison on the results of 1D and 2D best models applied to a detail of the painting *Putto*, Unknown author, 18th-century (circa) [1].

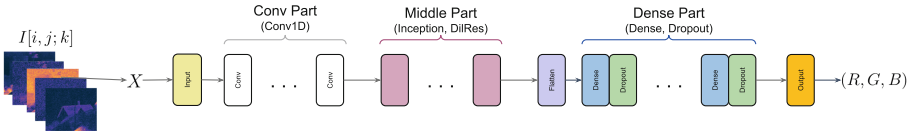

Fig. 3. Generic architecture for 1D models.

in particular, we exploited the Optuna ability of explore categorical parameter space to build different genres of layers (Fig. 3).

Note on the Loss Choice: we use the binary cross-entropy as loss, because is the same loss employed in [7].

3.1 1D Branch

For the 1D branch, we explore a variety of models' composition by join subsequent, different parts, each of which picked by optuna:

1. the first part may comprise various Conv1D networks, with various kernel sizes and number of filters, and may comprise a Batch Normalisation layer;
2. the second part is formed by a series of either Inception block or Dilated Residual block, their kernel size, and their number of filters. For the dilated residual blocks, their dilation rate (with dilation rate = 1 allowed, so we can have standard Residual block);
3. the third part is a series of couples Dense+Dropout; optuna picks the number of dense nodes, the dropout percentage and the activation function, chosen from the list ['RELU', 'SIGMOID', 'TANH', 'SOFTSIGN', 'SELU'].

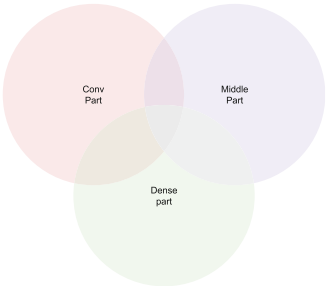

Fig. 4. All the 1D model possibilities explored by Optuna.

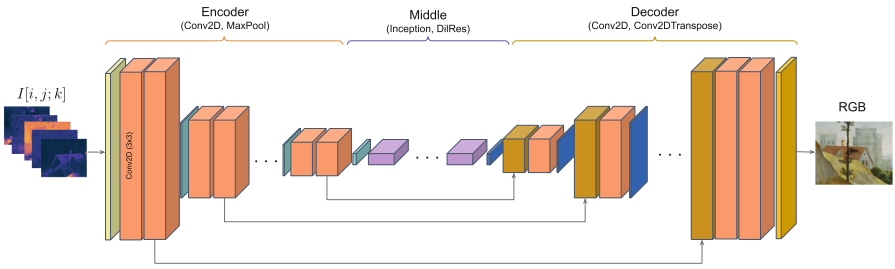

Fig. 5. Generic architecture for 2D models.

Notice that each part may have zero blocks; so, in one go, optuna is capable of explore either pure convolutional, pure dense, pure inception/resnet networks, and each possible 2- and 3-part combinations.

3.2 2D Branch

For the 2D branch, starting from a Encoder-Decoder UNet-shape of the model architecture (see Fig. 5), we let Optuna explore the possibilities, by either picking:

1. The number of Convolutional block layers in the Encoder/Decoder, and each of their kernel sizes and the number of filters;
2. The number of internal blocks;
3. For each internal block, its type (either Inception block or Dilated Residual block), its kernel size, and its number of filters. For the dilated residual block(s), its dilation rate (with dilation rate= 1 allowed, so we can have standard Residual block).

In this case, the Encoder convolutional blocks are based on the VGG-blocks, and can be either formed by 2- or 3- convolutional layers, followed by a $(2, 2)$ Max Pooling layer; similarly, the Decoder convolutional blocks can be either formed by 2- or 3- convolutional layers, preceded by a concatenating layer, merging the output of the previous block with the skip connection output coming from the

symmetric block, and followed by a transposed convolutional block for upsampling.

Technical Issue Faced: Unfortunately, due to hardware limitations (in particular, the limited size of the internal RAM of the GPU, and the in-ram initialisation size of the Optuna trials), we were not able to perform the training using the whole 500-channels energy depth, as in [7], where the $\mathcal{I}_{ij;k}$ tensors analysed have $(384, 288, 500)$ shape[3].

To reduce the in-memory size of the XRF tensors, we extracted the (supposedly) most relevant elemental maps:

Al (Ka), Si (Ka), K (Ka), Ca (Ka), Ti (Ka), Cr (Ka), Mn (Ka), Fe (Ka), Co (Ka), Ni (Ka), Cu (Ka), Zn (Ka), Ga (Ka), Se (Ka), Pd (Ka), Ag (Ka), Cd (Ka), Sn (Ka), Sb (Ka), Sb (La), W (La), Ir (La), Pt (La), Au (La), Hg (La), Pb (La).

This reduces the size of the tensor passed to the network to $(384, 288, 26)$, allowing the training. Nevertheless, we did not expect this to come without issues: sub-dominant peaks, as well as sub-dominant elements, may be as much important as those elements, and the coarse-graining of the input data may vastly impact the scores of the model, as we will see in the following section.

Table 4. Performances of the optimised models. The scores where the optimised model over-performs the respective best-models in [7] are written in red.

(a) Comparison table of the visual scores performances of the optimised models.

	SSIM	**MS-SSIM**	**PSNR**
1D	0.45916	0.73961	19.52408
2D	0.65938	0.64781	17.41894

(b) Comparison table of the losses performances of the optimised models.

1D	0.65856	0.01191
2D	0.49498	0.02302

4 Results and Discussion

The optimisation procedure was performed over 1000 trials; each training was performed over a 150-epochs, but with the caveat of invoking a callback for imposing an early stopping condition by monitoring the validation loss, to avoid overfitting. On top of that, Optuna imposes its pruning condition to abort unpromising trials.

[3] Which means each energy bin is $(38.5 - 0.5)\,\mathrm{KeV}/500 = 0.076\,\mathrm{KeV} = 76\,\mathrm{eV}$ wide.

4.1 1D Branch

The optimised model found by optuna is a model in the intersection of the Venn Fig. 4, since it has all three kinds of blocks:

- Convolutional Part:
  1. Input layer;
  2. Conv1D, $K = 6$ + ReLU activation;
  3. Conv1D, $K = 6$ + BatchNorm + ReLU activation;
  4. Conv1D, $K = 6$ + BatchNorm + ReLU activation;
  5. Conv1D, $K = 5$ + ReLU activation;
- Middle Part:
  1. Inception block;
  2. Inception block;
  3. Inception block;
  4. Inception block;
  5. Inception block;
  6. Inception block;
- Dense part:
  1. Average Pooling 1D, pool size = 6;
  2. Dropout layer, $\alpha = 0.43$;
  3. Flatten layers;
  4. Dense + selu activation;
  5. Dropout layer, $\alpha = 0.39$;
  6. Dense + sigmoid activation;
  7. Dropout layer, $\alpha = 0.23$;
  8. Dense + sigmoid activation;
  9. Output Dense layer;

where K is the (1-dimensional) kernel size, and *alpha* the drop-out rate.

Also, optuna picked a learning rate of 0.001639 for the RMSprop optimiser. The total number of parameters is 15.568.497, and the final binary cross-entropy loss was $\mathcal{L} = 0.5994$.

Once applied to the validation set, the results obtained are reported in Tables 4a, b, first row. We see that it slightly outperforms the best model of [7] in most scores, especially in the one found in that work as the most relevant, the MS-SSIM.

In Fig. 7 we report the visual confront between the best 1D model of [7] and the optimised model described above, applied on a validation set raw data. We can see a small improvement in the average appearance, especially on the dark tones, as the SSIM and MS-SSIM scores (reported in Table 5 tells us; nevertheless, it seems that, even if a little bit noisy, the model of [7] seems to be more capable of detecting rare pigment specimen, as the PSNR score seems to suggest (Fig. 6).

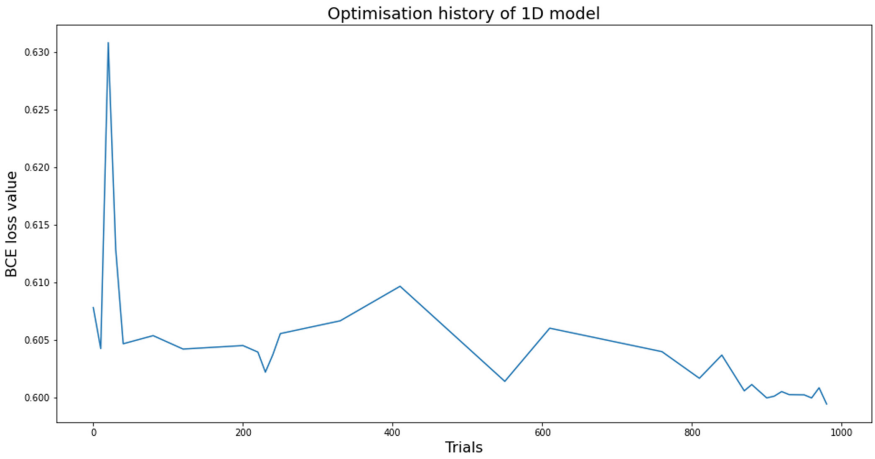

Fig. 6. History of optimisation for the 1D model.

4.2 2D Branch

The model found by optuna is:

- Encoder:
  1. Input layer;
  2. Conv Block, $K = (5, 5)$ + MaxPool $(2, 2)$;
  3. Conv Block, $K = (3, 3)$ + MaxPool $(2, 2)$;
  4. Conv Block, $K = (4, 4)$ + MaxPool $(2, 2)$;
  5. Conv Block, $K = (2, 2)$ + MaxPool $(2, 2)$;
  6. Conv Block, $K = (1, 1)$ + MaxPool $(2, 2)$;
- Middle Part:
  1. Inception block;
  2. Inception block;
  3. DilResNet block, $K = (4, 4)$, $d = 4$;
  4. DilResNet block, $K = (5, 5)$, $d = 2$;
  5. DilResNet block, $K = (5, 5)$, $d = 4$;
  6. DilResNet block, $K = (5, 5)$, $d = 4$;
- Decoder part [symmetric to Encoder) + output layer]:
  1. Conv Block, $K = (1, 1)$ + Conv2DT $(2, 2)$;
  2. Conv Block, $K = (2, 2)$ + Conv2DT $(2, 2)$;
  3. Conv Block, $K = (4, 4)$ + Conv2DT $(2, 2)$;
  4. Conv Block, $K = (3, 3)$ + Conv2DT $(2, 2)$;
  5. Conv Block, $K = (5, 5)$ + Conv2DT $(2, 2)$;
  6. Output Conv2D layer, $K = (3, 3)$;

where K is the (2-dimensional) kernel size, and d the dilation rate. Also, optuna picked a learning rate of 0.0004837 for the RMSprop optimiser. The total number of parameters is as small as 175.451, as somehow expected, due to the smallness

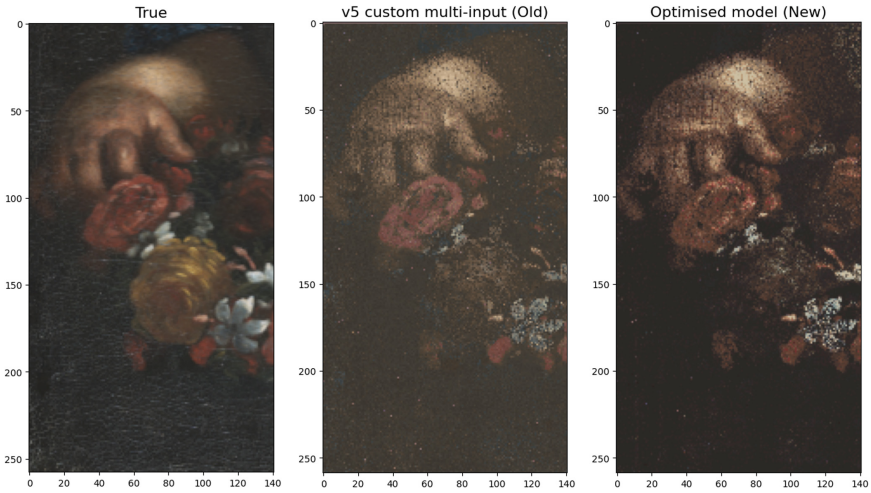

Fig. 7. Visual confront on the results of the 1D models; the true RGB image is shown on the left, the prediction of v5 custom multi-line model of [7] on the middle, and the prediction of the optimised model on the right.

of 2D dataset w.r.t. the 1D one, and the final binary cross-entropy loss was $\mathcal{L} = 0.39555$.

Unfortunately, due to the issues described in Sect. 3.2, the performances of the model, reported in the second row of Tables 4a, b, were slightly lower w.r.t. the ones of the best 2D model of [7]. This clearly implies that we can dramatically improve the hyperparameter optimisation of the 2D models, simply by improving the hardware set up.

Table 5. Comparison table of the visual scores performances of the 1D best old and new models on the detail of *Putto*, Unknown author, 18th-century (circa).

	SSIM	**MS-SSIM**	**PSNR**
1D v5_custom multi-input	0.497	0.840	22.048
1D Optimised	0.543	0.858	19.601

4.3 A Brief Weight Analysis

To try to shed light on the training process efficiency, we resort to the Theory of Implicit Heavy-Tailed Self-Regularization (HT-SR), in which a new Universal capacity control metric for Deep Neural Networks, α, is developed [20,22,23][4].

[4] For more theoretical details, see [21]. Link to paper.

HT-SR indicates that modern deep neural networks show a Heavy-Tailed Mechanistic Universality, meaning the spectral density of the normalised layer correlation matrix[5] can be fit to a power law, $\rho(\lambda) \sim \lambda^{-\alpha}$.

In [20,22,23], was shown that, empirically, a smaller value of α is correlated with better generalisation, with $\alpha \simeq 2$ universally across different best-in-class, pre-trained DNN architectures. So, it was empirically suggested that

- $\alpha < 2$: probably over-trained layer;
- $\alpha \simeq 2$: successfully-trained layer;
- $\alpha \gg 2$: poorly-trained layer.

In particular, the training of a layer is said to be

- Over-trained if $\alpha < 2$;
- Successfully trained if $2 \leq \alpha < 6$;
- Under-trained if $\alpha > 6$;

The implementation of these works is offered in the open source python package *WeightWatcher*[6]. The results are:

1. 1D:
   - average α: 2.00172
   - Percentage of over-trained layers: 33.33%
   - Percentage of successfully-trained layers: 66.66%
   - Percentage of under-trained layers: 0%
2. 2D:
   - average α: 6.09428
   - Percentage of over-trained layers: 8.62%
   - Percentage of successfully-trained layers: 51.72%
   - Percentage of under-trained layers: 39.66%

This seems to suggest that, as expected, the 1D network has experienced a more in-depth training (due to the size of its dataset), while the 2D network suffers the low-dimensionality of the dataset, and it was hard to properly train the model, to unleash its full capacity; this fact, joint to the technical issues reported in Sect. 3.2, may explain the non-completely satisfactory performances of the 2D models w.r.t. the 1D model. Nevertheless, it seems to suggest that we are set up for an improvement of the DNN performance by increasing the hardware resources, leaving room for further exploratory works.

[5] i.e., if $\mathbb{W}$ indicates the individual layer weight matrix, ρ is the density of the eigenvalues of

$$\mathbb{X} = \frac{1}{N} \mathbb{W}^T \mathbb{W}.$$

.

[6] GitHub: https://github.com/CalculatedContent/WeightWatcher.

5 Conclusions and Outlook

In this contribution, we outlined the results of hyperparameter optimisation of the model described firstly [7], developed withing the early stages of the AIRES-CH project, whose aim is to build a web-based app for digital restoration of pictorial artworks through Computer Vision technologies applied to physical imaging raw data.

Up to now, a set of 1D and 2D models to perform automatic recolouring were developed, trained, optimised and tested upon a relatively small dataset of X-Ray fluorescence imaging raw data, obtained from different kind of pictorial artworks, such as Flemish and modern multi-layered paintings, medieval illuminated manuscripts, and Renaissance drawings.

We have also employed these models in a web application for on-line analysis of XRF raw data, hosted in the CHNet cloud.

While the main result of [7] was to prove that the 1D and 2D branches are indeed able to provide complementary information to the reconstruction of the acquired image, here we tried to extend that result by means of hyperparameter optimisation, showing that, at least for the 1D branch, some improvement may be found.

Nevertheless, the main takeaway is that the small dimension of the dataset constitutes a limit to the learning capabilities of any network. Thus, to improve the scores of both branches by increasing the size of the dataset, a campaign of new XRF imaging measurements on pictorial artworks of the *Biblioteca Marucelliana* is scheduled to begin soon, in the context of the joint project AIRES-CH. Biblioteca Marucelliana will offer its vast collection of drawings, preparatory drawings, and sketches, all from the Italian *Renaissance*, ranging from the late XIV century to early XVI century; this will account for both an enlargement of the training dataset, as well as a standardisation of it, by having dataset entries of similar pictorial technique and created employing similar pigments; the analysis will also be conducted with the same apparatus (both in-situ as well as in-lab), providing a most valuable dataset for developing statistical analyses.

Furthermore, we have to develop the technique to take into account the presence of hidden pictorial layers and to recolour them, somehow factoring out the contribution from the outermost pictorial layer.

To do so, the full AIRES-CH DNN will comprise a *Generative, Adversarial* part [11]; the goal of this part is to associate a guessed XRF image to the colored image, which is the outermost pictorial layer; this is crucial to "extract" the covered image from the raw data. The generated XRF will have to be removed from the true raw data, to extract the hidden layer contribution to the XRF image.

This will finally allow to fed the generated XRF raw data to the recolouring architecture, to obtain a recoloured version of the *hidden* layers.

References

1. Ahmetovic, M.: Multi-analytical approach for the study of a XVII century Florentine painting: complementarity and data-crossing of the results of non-invasive diagnostics aimed at attribution and conservation. Master's thesis, University of Florence (2020)
2. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: a next-generation hyperparameter optimization framework (2019). https://doi.org/10.48550/ARXIV.1907.10902, https://arxiv.org/abs/1907.10902
3. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2019)
4. Albertin, F., et al.: "Ecce Homo" by Antonello da Messina, from non-invasive investigations to data fusion and dissemination. Sci. Rep. **11**(1), 15868 (2021). https://doi.org/10.1038/s41598-021-95212-2
5. Alfeld, M.: MA-XRF for historical paintings: state of the art and perspective. Microscopy Microanalysis **26**(S2), 72–75 (2020)
6. Bochicchio, L., et al.: Chapter 7 "Art is Not Science": a study of materials and techniques in five of Enrico Baj's nuclear paintings. In: Sgamellotti, A. (ed.) Science and Art: The Contemporary Painted Surface, pp. 139–168. The Royal Society of Chemistry (2020). https://doi.org/10.1039/9781788016384-00139
7. Bombini, A., Anderlini, L., dell'Agnello, L., Giacomini, F., Ruberto, C., Taccetti, F.: The AIRES-CH project: artificial Intelligence for digital REStoration of Cultural Heritages using physical imaging and multidimensional adversarial neural networks. Accepted for Publication on the ICIAP2021 Conference Proceedings (2021)
8. Bombini, A., et al.: CHNet cloud: an EOSC-based cloud for physical technologies applied to cultural heritages. In: GARR (ed.) Conferenza GARR 2021 - Sostenibile/Digitale. Dati e tecnologie per il futuro, vol. Selected Papers. Associazione Consortium GARR (2021). https://doi.org/10.26314/GARR-Conf21-proceedings-09
9. Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., James, S.: Machine learning for cultural heritage: a survey. Pattern Recogn. Lett. **133**, 102–108 (2020). https://doi.org/10.1016/j.patrec.2020.02.017
10. Gagliani, L.: Multi-technique investigations on a XIX century painting for the non-invasive characterization of visible and hidden materials and pictorial layers. Master's thesis, University of Florence (2020)
11. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016). http://www.deeplearningbook.org
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CoRR abs/1512.03385 (2015)
13. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Trans. Graph. **36**(4) (2017). https://doi.org/10.1145/3072959.3073659
14. Kleynhans, T., Schmidt Patterson, C.M., Dooley, K.A., Messinger, D.W., Delaney, J.K.: An alternative approach to mapping pigments in paintings with hyperspectral reflectance image cubes using artificial intelligence. Heritage Sci. **8**(1), 1–16 (2020). https://doi.org/10.1186/s40494-020-00427-7
15. Knoll, G.F.: Radiation Detection and Measurement, 4 edn. Wiley, Hoboken (2010)

16. Kogou, S., Lee, L., Shahtahmassebi, G., Liang, H.: A new approach to the interpretation of XRF spectral imaging data using neural networks. X-Ray Spectrometry **50**(4) (2020). https://doi.org/10.1002/xrs.3188
17. Larsson, G., Maire, M., Shakhnarovich, G.: FractalNet: ultra-deep neural networks without residuals. CoRR abs/1605.07648 (2016)
18. Licciardi, G.A., Del Frate, F.: Pixel unmixing in hyperspectral data by means of neural networks. IEEE Trans. Geosci. Remote Sens. **49**(11), 4163–4172 (2011). https://doi.org/10.1109/TGRS.2011.2160950
19. Mandò, P.A., Przybyłowicz, W.J.: Particle-Induced X-Ray Emission (PIXE), pp. 1–48. American Cancer Society (2016). https://doi.org/10.1002/9780470027318.a6210.pub3, https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470027318.a6210.pub3
20. Martin, C.H., Mahoney, M.W.: Heavy-tailed universality predicts trends in test accuracies for very large pre-trained deep neural networks (2019). https://doi.org/10.48550/ARXIV.1901.08278, https://arxiv.org/abs/1901.08278
21. Martin, C.H., Mahoney, M.W.: Universality and Capacity Metrics in Deep Neural Networks (2019)
22. Martin, C.H., Mahoney, M.W.: Implicit self-regularization in deep neural networks: evidence from random matrix theory and implications for learning. J. Mach. Learn. Res. **22**, 165:1–165:73 (2021)
23. Martin, C.H., Peng, T.S., Mahoney, M.W.: Predicting trends in the quality of state-of-the-art neural networks without access to training or testing data. Nat. Commun. **12**(1), 4122 (2021)
24. Mazzinghi, A., et al.: MA-XRF for the characterisation of the painting materials and technique of the entombment of christ by Rogier van der Weyden. Appl. Sci. **11**(13) (2021). https://doi.org/10.3390/app11136151
25. van den Oord, A., et al.: WaveNet: a generative model for raw audio. CoRR abs/1609.03499 (2016)
26. Ricciardi, P., Mazzinghi, A., Legnaioli, S., Ruberto, C., Castelli, L.: The choir books of San Giorgio Maggiore in Venice: results of in depth non-invasive analyses. Heritage **2**(2), 1684–1701 (2019). https://doi.org/10.3390/heritage2020103
27. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. CoRR abs/1505.04597 (2015)
28. Ruberto, C., et al.: La rete CHNet a servizio di Ottavio Leoni: la diagnostica per la comprensione dei materiali da disegno. In: Leo S. Olschki editore, F. (ed.) Accademia toscana di scienze e lettere la colombaria. atti e memorie, vol. LXXXV (2020)
29. Ruberto, C., et al.: Imaging study of Raffaello's La Muta by a portable XRF spectrometer. Microchem. J. **126**, 63–69 (2016). https://doi.org/10.1016/j.microc.2015.11.037
30. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2015)
31. Szegedy, C., et al.: Going deeper with convolutions. CoRR abs/1409.4842 (2014), http://arxiv.org/abs/1409.4842
32. Wang, M., Zhao, M., Chen, J., Rahardja, S.: Nonlinear unmixing of hyperspectral data via deep autoencoder networks. IEEE Geosci. Remote Sens. Lett. **16**(9), 1467–1471 (2019). https://doi.org/10.1109/LGRS.2019.2900733
33. Wang, Z., Simoncelli, E., Bovik, A.: Multiscale structural similarity for image quality assessment (2003). https://doi.org/10.1109/ACSSC.2003.1292216

34. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
35. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions (2016)
36. Zhang, X., Sun, Y., Zhang, J., Wu, P., Jiao, L.: Hyperspectral unmixing via deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett. **15**(11), 1755–1759 (2018). https://doi.org/10.1109/LGRS.2018.2857804
37. Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for neural networks for image processing. CoRR abs/1511.08861 (2015)