Multitask learning for denoising and analysis of X-ray polymer acquisitions

Nicolò Bonettini*, Carlo Andrea Gonano†, Paolo Bestagini*, Marco Marcon*, Bruno Garavelli†, Stefano Tubaro*
*Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano - Milan, Italy
†Xnext S.r.l. - Milan, Italy

***Abstract*—X-ray acquisitions are beneficial in food contaminant analysis as they can detect both metallic and non-metallic objects. This paper considers the scenario of single-pixel hyperspectral X-ray acquisitions applied to a series of materials with different characteristics. We propose a method that jointly applies a denoising operation and detects the analysed material in terms of a physical parameterisation. The proposed algorithm is based on a Convolutional Neural Network (CNN) trained with a multitask learning strategy using a custom loss function tailored to the problem at hand. Experimental results on metals and polymers show that the proposed method can also generalise to materials never seen at training time.**

***Index Terms*—X-ray imaging, CNN, polymer detection**

I. Introduction

Food safety is a discipline that oversees the preparation, handling and storage of food in order to prevent food-borne illness. Food contamination can happen during different stages of its preparation cycle: while cooking, during packaging, storing and transportation. For this reason, the European Commission has defined a series of food safety guidelines inspired by Hazard Analysis and Critical Control Points (HACCP) standards [1]. These impose standard food inspection rules and regulations on all producers.

There are different ways of detecting food contaminants. Some tests are intrusive and require physical access to the substance to be analysed. These include the vast majority of physical, chemical and biological laboratory analysis [2]. Other tests are non-intrusive and prove extremely important whenever physical access to food is not feasible, e.g., whenever food has already been packaged and sealed. This category of tests includes all kinds of scans performed while packages are transported by conveyor belts [3].

This paper focuses on a common category of industrial systems that play an essential role in food safety and quality assessment: X-ray sensors that capture hyperspectral images of objects running on a conveyor belt. Using X-rays has a series of advantages over other non-intrusive methods, e.g., metal detectors. X-rays can penetrate food products enabling accurate analysis of their interiors with no adverse effects [4]–[9]. Moreover, X-rays penetration depends on a series of factors: the atomic number, density and thickness of the elements composing the object under analysis [10], [11]. Therefore, X-ray hyperspectral images can be used as signatures of different materials. Additionally, if an object is a composite of different elements, it is possible to linearly combine the absorption effect of the different components [12]. Therefore, if we know the kind of food we are analysing and we want to detect physical contaminants within it, it is possible to subtract the specific food absorption signature from the performed X-ray acquisition, thus leaving only the possible contaminant absorption signature within the measurements [13]. This enables developing classification techniques that are agnostic to the kind of food a contaminant may be corrupting. In other words, recognising a contaminant alone helps us recognise the contaminant buried in food.

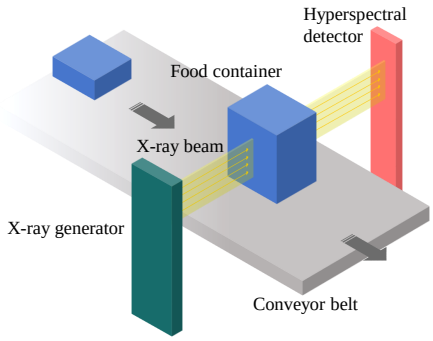

Fig. 1. Representation of the X-ray acquisition system for food contaminant classification.

In this paper, we consider the problem of joint denoising and analysis of hyper-spectral data acquired through an X-ray acquisition system. Given a single pixel of the used linear sensor, we propose a CNN that returns an estimate of a noise-free acquisition and detects the material (parameterized according to two physical quantities) scanned by the X-ray beam. First approaches to the use of CNNs for hyperspectral X-ray image denoising was proposed in [14], [15]. Unlike them, our method works on single pixels, rather than complete linear acquisitions, making the approach much more flexible and capable of detecting smaller contaminants. The proposed technique bases on a CNN, designed to achieve accurate performance while satisfying two application constraints: (i)

This research was partially supported by Xnext® s.r.l. who provided the complete data acquisition system together with insight and expertise that greatly assisted the research. Corresponding author email: nicolo.bonettini@polimi.it.

the analysis must run in real-time, thus the CNN must be fast enough when compared to the system acquisition rate (i.e., overly complicated architectures are not a choice); (ii) materials must be correctly recognized independently from their size as even tiny objects might be harmful in food safety. We test the proposed system over a dataset of 347 classes, obtained by combining 13 materials (polymers and metals) with different thicknesses. The experimental campaign is conducted to test the system performance even when some polymers have never been seen by the CNN during training. Results show that it is possible to estimate polymers' parameters even by using a single pixel of the studied sensor. Moreover, results suggest that the proposed CNN opens the doors to the possibility of synthetic data generation, which might be useful whenever acquisitions should be simulated.

II. Background and problem statement

This section introduces the reader to the considered X-ray acquisition system and Lambert-Beer's law, which explains the material parameterization used in this work. Finally, a formal definition of the problem faced in this paper is presented.

Acquisition system. The considered acquisition system is depicted in Fig. 1. It is composed of an X-ray generator, a conveyor belt carrying the object under analysis, and a hyperspectral X-ray linear detector. The generator emits the X radiation by directing a stream of high-speed electrons from the cathode to an X-ray vacuum tube's anode. The photons that the object has not absorbed reach the detector, which counts them at different spectral frequencies. In this framework, frequencies are commonly named X-ray energies due to the Planck-Einstein Relation [16].

The linear sensor samples the space in P different locations, thus can be considered as composed by P pixels. The acquired spectrum at each pixel location is divided into B photons frequencies intervals named "energy bins" from now on. For each time instant, the pixel detector acquires a B-element vector $\boldsymbol{y}$ containing the photons counts of all the energy bins.

Due to the stochastic photon emission, the acquired signal $\boldsymbol{y}$ is greatly affected by noise. In principle, we could approximate the clean acquisition, namely $\boldsymbol{\lambda}$, by averaging together an adequate number of the same object's acquisitions. Nonetheless, the system operates while the conveyor belt is moving, and the object intersects the X-ray stream just for a fraction of a second. Therefore, we cannot always use $\boldsymbol{\lambda}$ in our analysis, and it is mandatory to develop a time-efficient processing pipeline to work in such a real-time scenario.

Lambert-Beer law. Lambert-Beer's law relates the absorption of light to the material through which the light passes. Similarly, the absorption of X-rays at each energy E is related to the material through which the beam passes by the following equation:

$$\lambda_{\text{OUT}}(E) = e^{-\mu(E)\Delta x}\lambda_{\text{IN}}(E), \tag{1}$$

where $\lambda_{\text{IN}}(E)$ is the average number of incident X-ray photons with energy E, $\lambda_{\text{OUT}}(E)$ is the average number of transmitted ones, $\mu(E)$ is the linear attenuation coefficient of the material and Δx is the thickness of material through which X-rays have travelled. When an X-ray beam is acquired and digitalised by a sensor, energies E are discretised into B energy bins. As Lambert-Beer's law holds for every energy bin, we can rewrite it in vectorial form as:

$$\boldsymbol{\lambda}_{\text{OUT}} = \text{diag}[e^{-\mu_b \Delta x}]\boldsymbol{\lambda}_{\text{IN}}, \tag{2}$$

where μ_b is the linear attenuation coefficient for the b-th energy bin, diag is a $B \times B$ diagonal matrix, and both $\boldsymbol{\lambda}_{\text{OUT}}$ and $\boldsymbol{\lambda}_{\text{IN}}$ are B-element vectors. With this notation at hand, the sensor reading $\boldsymbol{y}$ is a measure of $\boldsymbol{\lambda}_{\text{OUT}}$.

The linear attenuation coefficient depends on the effective atomic number Z_{eff} of the material and the density-width $\rho\Delta x$ [12], [17]–[19]. This model allows us to bind each material to a pair of parameters.

Problem formulation. The problem considered in this paper is twofold. Given one noisy spectral acquisition $\boldsymbol{y}$ we are interested in: (i) estimate $\boldsymbol{\lambda}$, the clean version of it; (ii) estimate the two parameters Z_{eff} and $\rho\Delta x$ associated with the material. We consider the two estimations to be performed jointly as in a multi-task learning problem. Therefore, the goal of this paper is to design an operator $\Theta(\cdot)$ such that:

$$\hat{\boldsymbol{\lambda}}, (\hat{Z}_{\text{eff}}, \rho\hat{\Delta x}) = \Theta(\boldsymbol{y}), \tag{3}$$

where $\hat{\boldsymbol{\lambda}}$, $\hat{Z}_{\text{eff}}$ and $\rho\hat{\Delta x}$ represents the estimations of $\boldsymbol{\lambda}$, Z_{eff} and $\rho\Delta x$, respectively.

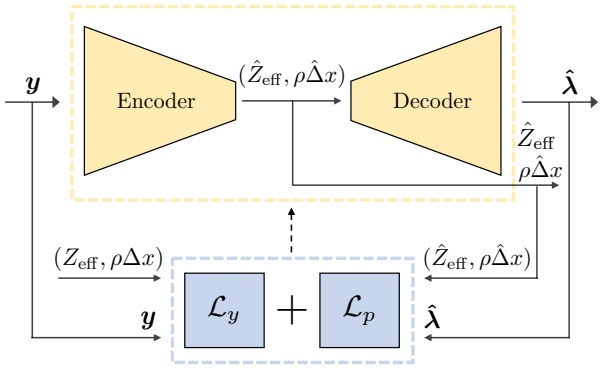

Fig. 2. Proposed network architecture and interaction between the CNN (yellow, top) and loss terms (blue, bottom).

III. Method

In our solution, we propose to use an autoencoder-like CNN. The network takes the spectrum under analysis as input, and it returns its denoised version and the two parameters Z_{eff} and $\rho\Delta x$ characteristic of the material under analysis. The network is trained according to a multi-task learning strategy. In the following, we report the details about the used architecture and the proposed custom loss function.

Network architecture. The proposed network architecture is depicted in Fig. 2. Following the classical Autoencoder paradigm, it is composed of an Encoder and a Decoder. The

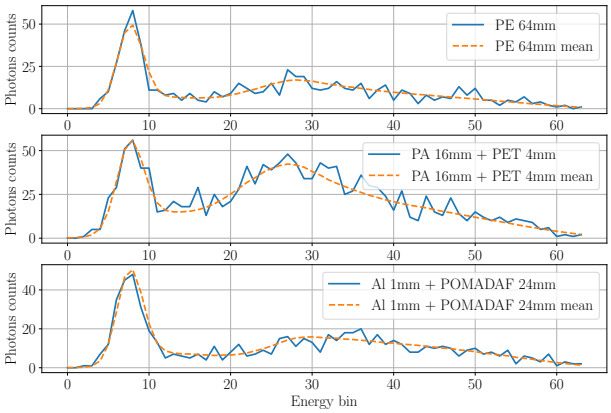

Fig. 3. Sample acquisitions from three different classes of the considered dataset. Blue curves show the actual noisy acquisition, whereas orange dashed curves show the clean ideal ground truth. Energy bins higher than $b = 64$ has been hidden for the sake of readability.

input is the spectral acquisition $\mathbf{y}$, whose dimensionality of B is reduced by the Encoder up to 2. The Encoder comprises 7 1D Convolutional Layers, each one followed by an ELU non-linearity layer [20]. Each layer's kernel size is used to reduce the signal's dimensionality through valid convolution, starting from a kernel size of 128 and halving it at each layer. The Decoder is designed with 7 1D Transposed Convolutional layers followed by ELUs to be symmetric to the Encoder and recover the original dimensionality B.

Loss function. The network is trained to minimise a compound loss function that takes into account two different kinds of errors: (i) the loss term $\mathcal{L}_y$ introduces a data fidelity term between the noisy input spectrum $\boldsymbol{y}$ and the reconstructed version of it $\hat{\boldsymbol{\lambda}}$; (ii) the loss term $\mathcal{L}_p$ takes the error on $\hat{Z}_{\text{eff}}$ and $\rho\hat{\Delta}x$ estimates into account. The generic form of the loss function is, therefore:

$$\mathcal{L} = \gamma\mathcal{L}_y + \mathcal{L}_p, \tag{4}$$

where γ is a weighting factor balancing the effect of one loss term over another.

We experimentally verified that the photons' count noise on each energy bin follows an independent Poissonian distribution. This allow us to employ a Weighted Log-Likelihood (WLL) as a data fidelity term $\mathcal{L}_y$:

$$\mathcal{L}_y = \frac{1}{B}\sum_{b=1}^{B}(\ln(P(y_b \mid \hat{\lambda}_b)) - \ln(P(y_b \mid y_b))), \tag{5}$$

where $P(y_b \mid \hat{\lambda}_b)$ represents the probability mass function for a Poisson distribution:

$$P(y_b \mid \hat{\lambda}_b) = \frac{\hat{\lambda}_b^{y_b} e^{-\hat{\lambda}_b}}{y_b!}. \tag{6}$$

and $P(y_b \mid y_b)$ is defined accordingly. By exploiting this prior, we do not need to know the actual clean ideal spectrum $\boldsymbol{\lambda}$ during training.

We also investigate the use of Mean Squared Error (MSE) as a term $\mathcal{L}_y$:

$$\mathcal{L}_y = \frac{1}{B}\sum_{b=1}^{B}\left(y_b - \hat{\lambda}_b\right)^2. \tag{7}$$

In this formulation, we are forcing the network to estimate the signal's clean version without explicitly inserting it in the equation. This is inspired by recent works on image denoising [21], [22], where the clean version of the input is unknown at training time.

As far as the term $\mathcal{L}_y$ is concerned, we exploit the Mean Absolute Percentage Error (MAPE) defined as

$$\mathcal{L}_p = \left|\frac{\hat{Z}_{\text{eff}}}{Z_{\text{eff}}} - 1\right| + \left|\frac{\rho\hat{\Delta}x}{\rho\Delta x} - 1\right|. \tag{8}$$

Notice that previous loss equations are defined for the i-th sample of the dataset. The global cost function on the whole dataset of N samples is defined as $\mathcal{L}_{tot} = \frac{1}{N}\sum_i^N \mathcal{L}_i$.

Deployment. Once the network is trained, we can feed it a spectrum $\boldsymbol{y}$ and obtain the clean version of the acquisition $\hat{\boldsymbol{\lambda}}$ alongside the estimation of the two parameters $\hat{Z}_{\text{eff}}$ and $\rho\hat{\Delta}x$ related to the material. We can also consider the Encoder and the Decoder part of the CNN separately. Using only the Encoder will allow us to estimate the parameters $\hat{Z}_{\text{eff}}$ and $\rho\hat{\Delta}x$ for detection purposes, whereas the Decoder could be used as a synthetic generator of spectra given two arbitrary parameters as input.

IV. Experiments and Results

This section describes the performed experimental campaign and reports the achieved results.

Experimental setup. To identify which combination of loss terms works best in the multi-task learning configuration and study the impact of the different loss terms separately, we devised a series of experiments with different network configurations.

To validate the multi-task pipeline, we consider two scenarios:

A) Multitask Autoencoder, with WLL as $\mathcal{L}_y$ and MAPE as $\mathcal{L}_p$.
B) Multitask Autoencoder, with MSE as $\mathcal{L}_y$ and MAPE as $\mathcal{L}_p$.

In these two methods, the network input is the noisy spectral acquisition $\boldsymbol{y}$. The network outputs are the estimated clean spectrum $\hat{\boldsymbol{\lambda}}$ and the estimated parameters $\hat{Z}_{\text{eff}}$ and $\rho\hat{\Delta}x$ from the latent space of the network.

To study the effect of the single loss terms, we consider three scenarios:

C) Only Encoder, with MAPE as $\mathcal{L}_p$.
D) Full Autoencoder, with WLL as $\mathcal{L}_y$.
E) Full Autoencoder, with MSE as $\mathcal{L}_y$.

In scenario C, we consider only the encoding part of the network. The input is the noisy spectral acquisition $\boldsymbol{y}$, while the outputs are directly the latent space parameters $\hat{Z}_{\text{eff}}$ and

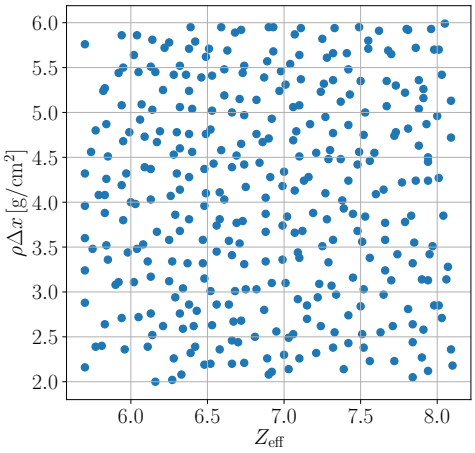

Fig. 4. Z_{eff} and $\rho\Delta x$ true distributions observed in the adopted dataset.

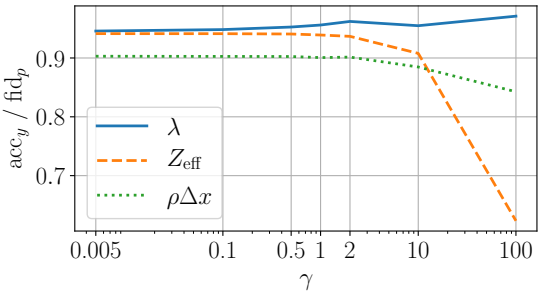

Fig. 5. Performances of $\hat{\boldsymbol{\lambda}}$, $\hat{Z}_{\text{eff}}$ and $\rho\hat{\Delta}x$ varying γ.

$\rho\hat{\Delta}x$. In methods D and E, we do not consider the multi-task learning part, feeding the network $\boldsymbol{y}$ and obtaining $\hat{\boldsymbol{\lambda}}$ without extracting the latent space parameters.

As stated in Sec. II, we obtain $\boldsymbol{\lambda}$ by averaging 2 560 acquisition together. It is worth noting that excluding method C, we train all the other methods to reconstruct $\hat{\boldsymbol{\lambda}}$ (the mean version of the spectrum). However, the loss function $\mathcal{L}_y$ takes as input the noisy acquisition $\boldsymbol{y}$, and we never use $\boldsymbol{\lambda}$ during training. This choice is strictly related to the exploitation of the inherent prior introduced by $\mathcal{L}_y$, which allows the network to estimate the mean spectrum $\boldsymbol{\lambda}$ even without seeing one.

At test time, we are interested in measuring network performance. Therefore, we derive two metrics based on the adopted loss functions. In case of single parameters Z_{eff} and $\rho\Delta x$, we simply use $1-$MAPE to obtain a fidelity index $\text{fid}_p \in (-\infty, 1]$, the higher the better. In the case of spectrum, we derive a compatibility index $\text{acc}_y \in [0, 1]$ by computing:

$$J = -\frac{1}{B}\sum_{b=1}^{B}(\ln(P(\lambda_b \mid \hat{\lambda}_b)) - \ln(P(\lambda_b \mid \lambda_b))), \quad (9)$$

$$\text{acc}_y = \mathrm{e}^{-J}, \quad (10)$$

that is computing Eq. (5) using the clean version of the spectrum $\boldsymbol{\lambda}$ instead of the noisy acquisition $\boldsymbol{y}$ and taking the negative exponential. When equal to 1, acc_y indicates a perfect match between the two spectra, when equal to 0 it indicates a miserable match.

Regarding the implementation, we resort to Pytorch framework for developing the CNN. All the experiments are run on a workstation equipped with an NVIDIA Titan V Graphics Processing Unit (GPU), an Intel Xeon E5-2687W and 256 GiB RAM.

Dataset. The dataset is composed of acquisitions taken with the system mentioned above. The operating point of the X-ray tube is 60keV and 0.3 mA. We place each material object in front of the sensor and irradiate it with an X-Ray beam, recording the linear sensor output. For each object, we acquire 2 560 static images, with acquisition time 2ms (i.e., the active recording time of the sensor). The frequency spectrum of the acquisition is divided into $B = 256$ energy bins. We consider only the central pixel of the sensor. Therefore, each acquisition's final shape is (1×256) where the first dimension represents the pixel, and the second dimension is the number of energy bins used to represent the spectrum. Fig. 3 shows three sample acquisitions, each belonging to a different class (i.e., combinations of material and thickness).

We consider 12 different polymers: 'PE', 'PA', 'PC', 'PMMA', 'PET', 'POM', 'POMADAF', 'PVDF', 'PTFE', 'PBT', 'PPS', 'PVC' and the metal Aluminium. These represent some of the most common contaminants that can be found during food inspection. Acquisitions are made both on single materials and on pairs of tiled materials. Considering all combinations of materials and thicknesses, we end up with 347 different classes, with a roughly uniform distribution of parameters Z_{eff} and $\rho\Delta x$, as shown in Fig. 4.

The whole corpus of images is composed of more than 908 445 acquisitions of shape (1×256). We remove all the acquisitions of materials 'PA' and 'POMADAF' from the whole corpus. We also remove all the acquisitions of all the other materials at thickness 24mm and 36mm. From this reduced corpus, we use the 70% to create the train set $\mathcal{D}$, and the remaining 30% composes the first test set $\mathcal{T}$. We use an additional 20% of the train set as a validation set to evaluate the training procedure. The acquisitions of materials 'PA' and 'POMADAF' compose the second test set $\mathcal{T}_m$, while the acquisitions of all the materials with thickness 24mm and 36mm compose the third test set $\mathcal{T}_t$. The purpose of the last two test sets is to study the system behaviour with unknown (i.e., previously unseen) classes (i.e., materials and thicknesses). Those class pairs are randomly chosen.

Parameter search. Since the loss function is a weighted average between two terms, we are interested in finding the best weighting factor γ. Therefore, we run a preliminary set of experiments on $\mathcal{T}$ varying γ. Results reported in Fig. 5 shows how incrementing γ gives a little benefit in estimating $\hat{\boldsymbol{\lambda}}$, while affecting the estimation of $\hat{Z}_{\text{eff}}$ and $\rho\hat{\Delta}x$. To obtain a balance between the three estimations, we fixed $\gamma = 0.1$.

Results on known dataset $\mathcal{T}$. We report results for the known dataset in Table I. Regarding $\boldsymbol{\lambda}$, the multitask approach A offers a better performance than its single task counterpart E,

TABLE I
RESULTS ON KNOWN DATASET $\mathcal{T}$.

Method	Z_{eff} (fid_p)	$\rho\Delta x$ (fid_p)	$\boldsymbol{\lambda}$ (acc_y)
MAPE + MSE (multitask) (A)	0.94227	0.90416	0.97946
MAPE + Poisson (multitask) (B)	0.93868	0.90469	0.96163
MAPE ($\rho\Delta x$, Z_{eff}) (C)	0.94140	0.90355	-
MSE ($\boldsymbol{\lambda}$) (E)	-	-	0.93835
Poisson ($\boldsymbol{\lambda}$) (D)	-	-	0.96395

TABLE II
RESULTS ON UNKNOWN DATASET $\mathcal{T}_m$ (UNKNOWN MATERIAL).

Method	Z_{eff} (fid_p)	$\rho\Delta x$ (fid_p)	$\boldsymbol{\lambda}$ (acc_y)
MAPE + MSE (multitask)	0.93758	0.90864	0.97864
MAPE + Poisson (multitask)	0.93462	0.90947	0.95950
MAPE ($\rho\Delta x$, Z_{eff})	0.93659	0.90817	-
MSE ($\boldsymbol{\lambda}$)	-	-	0.93613
Poisson ($\boldsymbol{\lambda}$)	-	-	0.96267

while B is comparable with D. The performance on Z_{eff} and $\rho\Delta x$ is pretty much the same when comparing with method C. Probably, the estimation of the two parameters is a more straightforward task for the network, and it does not need any help from the multitask learning paradigm. Conversely, the clean signal estimation is a more challenging task and benefits from the joint loss function.

Results on unknown datasets $\mathcal{T}_m$ and $\mathcal{T}_t$. Table II and Table III show the results when we test on unknown materials and thicknesses respectively. We can confirm the trend of Table I: in both cases, the network reaches the best accuracy on estimating $\boldsymbol{\lambda}$ when training in a multitask fashion. The estimation of Z_{eff} and $\rho\Delta x$ seems to suffer less from the lack of multitasking.

V. CONCLUSIONS

This paper proposes a method for jointly denoising a hyperspectral X-ray acquisition and estimating the two physical parameters related to the acquired object. We tested our approach to a vast dataset of polymers and metal acquisitions, considering the case in which the network has never seen the material under analysis. We compared the multitask paradigm with the separate estimation of the two desired outputs (e.g., the parameters and the mean spectrum), showing benefits in adopting the former. Future work will focus on applying the decoding part of the network as a physical generator of synthetic data.

TABLE III
RESULTS ON UNKNOWN DATASET $\mathcal{T}_t$ (UNKNOWN THICKNESS).

Method	Z_{eff} (fid_p)	$\rho\Delta x$ (fid_p)	$\boldsymbol{\lambda}$ (acc_y)
MAPE + MSE (multitask)	0.94062	0.90428	0.97930
MAPE + Poisson (multitask)	0.93676	0.90276	0.96203
MAPE ($\rho\Delta x$, Z_{eff})	0.93976	0.90394	-
MSE ($\boldsymbol{\lambda}$)	-	-	0.93288
Poisson ($\boldsymbol{\lambda}$)	-	-	0.96506

REFERENCES

[1] Norman Marriott and Robert Gravani, "The role of HACCP in sanitation," in *Principles of Food Sanitation. Food Science Texts Series*, pp. 99–115. Springer, Boston, MA, 2006.

[2] Food and Agriculture Organization of the United Nations., *Risk-based food inspection manual*, Food and Agriculture Organization of the United Nations, Rome, Italy, 2008.

[3] Shyam N Jha, *Nondestructive evaluation of food quality: theory and practice*, Springer, 2010.

[4] Sunil Mathanker, Paul Weckler, and Tim Bowser, "X-ray applications in food and agriculture: a review," *Transactions of the ASABE (American Society of Agricultural and Biological Engineers)*, vol. 56, pp. 1227–1239, 05 2013.

[5] R. A. Speir and M. A. Haidekker, "Onion postharvest quality assessment with x-ray computed tomography – a pilot study," *IEEE Instrumentation Measurement Magazine*, vol. 20, no. 3, pp. 15–19, June 2017.

[6] I. Khosa and E. Pasero, "Pine nuts selection using x-ray images and logistic regression," in *2014 World Symposium on Computer Applications Research (WSCAR)*, Jan 2014, pp. 1–5.

[7] I. Khosa and E. Pasero, "Feature extraction in x-ray images for hazelnuts classification," in *2014 International Joint Conference on Neural Networks (IJCNN)*, July 2014, pp. 2354–2360.

[8] I. Khosa and E. Pasero, "Defect detection in food ingredients using multilayer perceptron neural network," in *2014 World Symposium on Computer Applications Research (WSCAR)*, Jan 2014, pp. 1–5.

[9] Q. Lu, F. Wang, Y. Li, and J. Cai, "Real-time nondestructive inspection of chestnuts using x-ray imaging and dynamic threshold," in *2010 World Automation Congress*, Sep. 2010, pp. 365–368.

[10] A. D. Mc Naught and A. Wilkinson, *IUPAC Compendium of Chemical Terminology – The Gold Book*, Blackwell Scientific Publications, Oxford, 1997.

[11] M. Wocher, K. Berger, M. Danner, W. Mauser, and T. Hank, "Hyperspectral retrieval of canopy water content through inversion of the beer-lambert law," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, July 2018, pp. 3805–3808.

[12] Shivangi Kelkar, Carol J. Boushey, and Martin Okos, "A method to determine the density of foods using x-ray imaging," *Journal of food engineering*, vol. 159, pp. 36–41, 2015.

[13] George Wypych, "Photophysics," in *Handbook of Material Weathering*, pp. 1–26. Elsevier, 2018.

[14] M. Re, "Denoising and classification of hyperspectral X-ray images for food quality assessment," *POLITesi - Politecnico di Milano*, 2018.

[15] N. Bonettini, M. Paracchini, P. Bestagini, M. Marcon, and S. Tubaro, "Hyperspectral X-ray denoising: model-based and data-driven solutions," in *European Signal Processing Conference (EUSIPCO)*, 2019.

[16] David J. Griffiths, *Introduction to Quantum Mechanics*, Prentice Hall, 1994.

[17] F. W. Spiers, "Effective atomic number and energy absorption in tissues," *The British Journal of Radiology*, vol. 19, no. 218, pp. 52–63, 1946.

[18] R C Murty, "Effective atomic numbers of heterogeneous materials," *Nature*, vol. 207, 1965.

[19] G. Bubba, "Material identification and Compton scattering effects in X-ray multienergy radiography of homogeneous compounds," *POLITesi - Politecnico di Milano*, 2017.

[20] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," in *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun, Eds., 2016.

[21] Allard A. Hendriksen, Daniel M. Pelt, and K. Joost Batenburg, "Noise2Inverse: Self-supervised deep convolutional denoising for tomography," *arXiv e-prints*, Jan. 2020.

[22] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila, "Noise2Noise: Learning image restoration without clean data," in *Proceedings of the 35th International Conference on Machine Learning*, Jennifer Dy and Andreas Krause, Eds., Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018, vol. 80 of *Proceedings of Machine Learning Research*, pp. 2965–2974, PMLR.