IFC-Bank of Italy Workshop on "Data science in central banking: enhancing the access to and sharing of data"

17-19 October 2023

Siamese neural networks for detecting banknote printing defects[1]

Katia Boria, Andrea Luciani, Sabina Marchetti and Marco Viticoli,
Bank of Italy

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

Siamese neural networks for detecting banknote printing defects

K. Boria, A. Luciani, S. Marchetti and M. Viticoli [1]

Abstract

The production of banknotes is a complex process, composed of different printing steps, in which various kinds of defects can be generated that, if not adequately monitored, can lead to production waste, significantly impacting productivity and costs. This paper proposes a new approach for identifying defects during banknote production using *one-shot learning* methods. These methods rely on a small number of observations in order to train a Siamese neural network to reproduce the similarities between pairs of samples. The network can then identify defects in new banknote images by comparing them to benchmark samples. The proposed approach allows the correct identification of some specific defects on banknotes, even with limited training data, laying the foundation for the development of a solution for recognition and intelligent classification of defects on banknotes.

Keywords: banknotes production, artificial intelligence, neural networks, one-shot learning, quality control

JEL classification: C45, L15, L69, O39

Contents

[1] Bank of Italy. The views expressed in this article are those of the authors and do not necessarily reflect those of the Bank of Italy.

1. Introduction

Similar to other industries, banknote production relies on a comprehensive system of quality controls. In the Eurosystem, such controls are conducted throughout the entire production process (raw materials, semi-finished and finished banknotes) in compliance with the relevant ISO standards and through rigorous procedures defined by the European Central Bank (ECB) and harmonized among the various printing works to ensure a high degree of homogeneity of euro banknotes. With particular reference to quality checks on finished banknotes, the number and type of printing defects acts as a key discriminating parameter for the conformity of each production batch. Today, the assessment of defects is carried out by operators through visual examination of the banknotes. Therefore, although there is an articulated set of safeguards aimed at ensuring the application of objective criteria, the experience and sensitivity of the workers play a decisive role in the evaluation process, which could therefore be affected by subjectivity factors.

In principle, automating quality control of industrial products helps make processes more accurate and efficient. Artificial intelligence, and in particular neural networks in deep learning, are useful in this scenario due to their ability to represent and describe systems characterized by high complexity and variability. At the printing stage, the evaluation and identification of potential defects by neural networks enables a reduction in operational costs by devolving to quality control specialists the detailed analysis of potential defects found and their causes.

This paper reports on what has been observed as part of an exploratory study on the use of neural networks for automated recognition and classification of a particular type of banknote defects in the context of quality control. Section 2 provides an essential overview of the quality control process for detecting print defects on banknotes. Section 3 outlines the main technical features of Siamese neural networks - that is, the class of deep learning models used in the present study - as well as the set of input images used for training and calibration. The main aspects of model training and the results achieved using Siamese networks are reported and discussed in Section 4. Finally, Section 5 outlines possible directions for methodological and operational research and development.

2. Defects in banknote printing and production variability

As part of the production process, strict quality controls are carried out on the produced and semi-finished banknotes, in compliance with the European Central Bank's (ECB) quality benchmark decisions and the requirements of the Integrated Quality, Environment and Occupational Health and Safety Management System.

Together, these checks provide important information about potential process criticalities and final product defects; this information can be profitably used to intervene in the process itself to reduce production waste in the printing process and even marginal defects in the finished product.

The verification of many parameters has already been automated through the introduction of optical measuring systems. However, visual acceptance checks on finished banknotes are still assigned to highly trained staff, who perform these tasks manually on a representative sample of each production batch. Defects are then identified and classified based on ECB documents and reference samples. Classification is done by defect type and size, as well as by area of the surface where different defects may be found. The final conformity assessment of the production batch then depends on the number and classification of defects found on the representative sample of the batch. This process is based on a set of rules and comparison with reference samples, similar to what is performed automatically by *deep learning* techniques.

For the experimentation described in this paper, a pilot application was chosen for the detection of defects on the €50 banknotes of the Europa series[2] currently in production at the Bank of Italy. In this first phase, the study focused on a specific type of defect present on the European flag, the so-called *bite* defect, which consists of a lack of ink on a homogeneous background, and can vary in shape, size and position. The choice was determined by the relevance and number of occurrences referring to this defect type in the production batches examined at the time of experimentation.

The variability of the banknotes was another important factor of analysis that had to be taken into account in the experimental phase. Since banknotes result from different printing stages, they may exhibit appreciable differences between them that reflect normal production variability and hence should not be considered defects. In particular:

1. the position of elements on the banknote surface may vary within fixed margins of tolerance;

2. the density and colour of inks can vary, resulting in different shades and intensities;

3. the acquisition phase, either manual or automatic, can introduce additional image variability.

[2] The Europa series is the second series of Euro banknotes, introduced from 2013 to 2019.

In addition to product variability, it was also necessary to account for the variability of defects to be identified and classified, differing in size, location and colour intensity.

We initially proceeded with manual acquisition of high-resolution (~2656x1467 pixels, 24-bit colour, RGB model) images of each banknote, with the aim to minimize the probability of acquisition defects. In order to reach a sufficient number of samples to be processed, we also considered images acquired automatically during the production process, which had a lower resolution (~675x388 pixels, 24-bit colour, RGB pattern). For both cases, the acquisition of banknote images required operator intervention, although in the second case the procedure was significantly faster. Manual acquisition, on the other hand, allowed for a more careful selection of the defects to be acquired, representative of a longer production time frame (several weeks) and greater production variability. At present, the two different acquisition approaches are to be taken into account when training the network for their effects on the availability, extent and variability of the image sample.

3. Methodology

The proposed application focuses on the detection of bite-type printing defects using Siamese neural networks. This section introduces the reader to the technical aspects of the analysis conducted, which include: i) the type of data representation of digital images, i.e., ordered sets of matrices called *tensors*; ii) the convolutional neural networks used for tensor processing, and the limitations of such models in this specific application of banknote images elaboration; iii) the specialization of deep learning, known as one-shot learning, and the related class of Siamese neural networks.

3.1 Data

From a technical point of view, the images belong to the type of data in "unstructured" format. For the purpose of statistical learning, each image is mathematically represented by an ordered set of three matrices, or three-rank tensor, which captures its relevant characteristics. Specifically, the so-called pixels[3], i.e. the elementary graphic components of each image, are mapped into groups of elements of the matrices. Within a tensor, each element of a matrix assigns a numerical value to its corresponding pixel; in the case of 24-bit colour, values vary between 0 and 255 and indicate the intensity of the colour associated with the matrix according to its placement in the sequence. In the case of the RGB model, the first matrix represents the so-called Red colour channel (*R*), the second matrix represents the Green colour channel (*G*), and the third represents the Blue

[3] Pixels are minimal units of the surface area of a digital image.

colour channel (*B*)[4]. As an example, according to such a representation scheme, a blue pixel in the image is obtained by setting to zero the related matrix values in the R and G channels, while setting to 255 the matrix value in the R channel. In the general case, given an image resolution of NxM pixels encoded via the RGB model, the image will be numerically represented by a three-rank tensor containing NxMx3 integers, each of which can as mentioned vary between 0 and 255.

3.2 Deep learning and convolutional neural networks

Advanced image processing generally makes use of artificial intelligence techniques, particularly deep learning, for automatic recognition or classification of objects represented in an image. Deep learning is a specialization of *machine learning*, extending the process of learning a representation-rule mapping inputs to outputs to sequences of representation-layers (Goodfellow, Bengio & Courville, 2016). The approach is based on the *data-driven* paradigm, according to which the information necessary for the understanding of a given phenomenon. The proper functioning of the related representation model[5] is learned from a set of empirical data, called the *training* set. By design, training of complex machine learning model requires the specification and formalization of a limited number of assumptions about the phenomenon being analysed.

Neural networks are models for deep learning equipped with internal processing components (layers) iteratively delivering a representation of inputs. Within each layers, information processing is based on units called *neurons*. In their simplest formalization, so-called "fully connected", layers are composed of neurons that receive, process, and transmit all information coming from a preceding layer to the next.

In a fully connected neural network, input observation $\boldsymbol{y}^0 \in \mathbb{R}^{n_0}$ is sequentially mapped into projection $\boldsymbol{y}^{l+1} \in \mathbb{R}^{n_{l+1}}$, the projection of the input observation produced by the l-th fully connected layer, with $l=0,\ldots,L$, derived through the recursive formula:

$$\boldsymbol{y}^{l+1} = g^l(\boldsymbol{w}^l \boldsymbol{y}^l + \boldsymbol{b}^l).$$

g^l is a generic (possibly non linear) *activation function* that applies to the linear transformation of the vector $\boldsymbol{y}^l \in \mathbb{R}^{n_l}$. Parameters' values, i.e. the elements from matrix of weights $\boldsymbol{w}^l \in \mathbb{R}^{n_{l+1} \times n_l}$ and from vector of bias terms $\boldsymbol{b}^l \in \mathbb{R}^{n_{l+1}}$, are defined by the learning process, provided a random initialization, based on the training set.

The training of a model is also subject to a "calibration" phase, to validate its performance on an unseen set of observations and to establish optimal values of the model's hyper-parameters, i.e. its architecture - defined by the type of layers

[4] Another frequently used model is the so-called HSV, for which cell values vary between 0 and 360. The first channel corresponds to the hue (*Hue*). The second channel identifies the degree of saturation (*Saturation*). The third channel refers to the brightness of the colour (*Colour Value*).

[5] The proper functioning of the model is expressed by the value of an objective function, which guides its training process.

that regulate transmission of information; the depth of the neural network, i.e., the number of hidden layers; the number of neurons per layer and the activation function - and features of the training optimization routine, including the learning algorithm.

Image processing applications typically make use of neural networks containing convolutional layers (*Convolutional Neural Networks*, CNNs)[6]. CNNs are inspired by the biological mechanism of visual perception. They operate a reduction in the complexity of the input data through targeted extraction of a synthesis to make processing efficient. Technically, convolutional layers process the information represented by a tensor according to a procedure based on recursive filters that slide along surface dimensions, that is, throughout groups of pixels adjacent to each other, or region. From each region, projections called *filter maps* are extracted.

In order to ensure the stability of the information processing system, convolutional layers alternate with *pooling* layers that synthesize and aggregate the contribution of several adjacent regions into a filter map (Boureau *et al.,* 2010). Typically, pooling layers collect either the average value (*average pooling*) or the maximum value (*max pooling*) of adjacent regions from each region. Pooling usually results into a lower-dimension tensor whose elements will serve as input for the next layer.

By construction, the recursive region-based nature of the detection scheme entails replication of information from same pixel across multiple projections. As a consequence, each filter map is sensitive to small shifts of an element within the image area, and thus to small shifts of the pixels that contribute to define it. This is particularly relevant in the use case considered, since banknotes are characterized by an established organization of the elements represented on their surface, although with tolerated variability given by their reciprocal positioning and colour intensity. The use of CNN for classification in image processing can be distinguished into approaches based i) on input segmentation and recognition of different elements (*object recognition*) or ii) on labelling an input image (*image recognition*). Object recognition can either involve image segmentation into specific areas of the surface identified as potentially relevant, to be passed on to ad hoc image recognition models for labeling individual features, or incorporate the two phases within a single neural network (Redmon *et al.,* 2016). CNNs for recognizing different elements of an image are usually characterized by complex architectures, which are associated with significant computational costs for the training and application phases. The literature related to image classification has over time produced CNNs characterized by high performance on large volumes of data, partly due to the refined complexity of the architectures (Krizhevsky *et al.,* 2017). In particular, the accuracy of the classification process is enabled by deep neural networks organized according to hierarchical structures (Simonyan & Zisserman, 2014; Szegedy *et al.,* 2015; Yan *et al.,* 2015) and granular parameterization of layers, in which the size of recursive filters approximates the single pixel (Zeiler & Fergus, 2014).

[6] LeCun *et al.* (1989).

The training process of a CNN consists in the global optimization of the model, through the progressive adaptation of its parameters' value. To handle high dimensionality of the input and the recursive nature of its processing, CNNs used for image recognition are characterized by a significant number of parameters to be estimated, and thus require large training sets to converge. Since acquisition of a large number of training sample is especially costly in our setup, sound training of CNNs represents a challenging task.

Flexibility of models characterized by complex architectures is known to expose classification exercises to so-called *overfitting*. Overfitting consists in the modeling of irreducible error in the training data, that penalizes the model's ability to generalize the prediction performance. To mitigate this risk and enhance convergence of the training process, while addressing the computational complexity of processing, a number of steps can be taken. Our application, in particular, makes use of: i) ad hoc methods for parameters initialization, assigning initial values within optimal ranges (Domínguez, 2020), such as the so-called *Xavier* method for defining the parameters of a uniform distribution (Glorot & Bengio, 2010); ii) optimal calibration of the training algorithm, typically belonging to the *Stochastic Gradient Descent* (SGD) category for CNNs (Qian, 1999; Wijnhoven & De With, 2010); iii) input normalization (via *batch normalization*) to smooth the values of the objective function and its gradient derivative (Ioffe & Szegedy, 2015); iv) adoption of regularization tools such as dropout (Hinton *et al.,* 2012), i.e., random switch-off of links between neurons during training, and quadratic regularization of weights in convolutional layers (Yu *et al.,* 2008); v) re-initialization of the value of parameters when local optima are achieved, to improve exploration of the solution space (Treadgold & Gedeon, 1996; Guo & Li, 2006).

3.3 Convolutional neural networks for the study of banknotes

CNNs can improve banknote image analysis mainly due to two factors: the flexibility of the models and their ability to process data of digital image without resorting to intermediate tools to extract characteristic information, or *features* (Lee *et al.,* 2017).

In the literature addressing processing of banknote images, applications typically make use of statistical-mathematical models and machine learning techniques to recognize denomination (Grijalva *et al.,* 2010; Sharma *et al.,* 2012; García-Lamont *et al.,* 2013), serial number (Feng et al, 2014; Liu *et al.,* 2010; Wenhong *et al.,* 2010; Hasanuzzaman *et al.,* 2011), currency (Manikandan & Sumithra, 2015), wearing (Sun & Li, 2008; Daraee & Mozaffari, 2010; Mousavi *et al.,* 2015) or for the detection of counterfeit specimens (Darade *et al.,* 2016; Suresh *et al.,* 2016). Advanced analytical techniques are also employed for counting, operational support to visually impaired individuals, and quality monitoring of circulating banknotes among individuals, merchants and banks themselves.

Recent literature also highlights publications using CNNs for banknote recognition (Jadhav *et al.,* 2019; Zhang *et al.,* 2019; Jang *et al.,* 2020; Park *et al.,* 2020; Veeramsetty *et al.,* 2020) and counterfeit specimen detection (Kamble *et al.,* 2019; Sawant *et al.,* 2022), achieving an overall increase in performance. In order to overcome the challenges associated with the use of CNNs, such as the high

volume of observations requested for training, and the tendency of CNNs to overfitting given the variability characteristics associated with banknotes, CNNs have in some cases been developed according to *adversarial learning* approaches for generating synthetic images (Ali *et al,* 2019; Desai et al., 2021; Khemiri *et al.,* 2022) or *transfer learning*, i.e., approaches that make use of already trained models and then operate their specialization, or fine-tuning, through a less intensive training process (Laavanya & Vijayaraghavan, 2019; Linkon *et al.,* 2020; Aseffa *et al.,* 2022).

Applications referring to banknote production can also be found in literature (Ke *et al.,* 2016; Pham *et al.,* 2017). In particular, research by Ke and co-authors focuses on the detection of bite-like defects - the same class of defects used in this work - by developing a CNN. Architectural details of the network and information on the training set, however, are not fully exposed by the authors.

The review of methodologies proposed in the literature for image processing of banknotes through deep learning techniques underscored that, for the purposes of our application, training a sufficiently deep CNN would have a number of limitations, given by the aforementioned variability in the arrangement of graphical elements, defects on the image surface, and the limited availability of observations for training. To mitigate the relatively limited availability of observations, a preliminary extension of the training set was considered, by resorting to *data augmentation* solutions (Wang & Perez, 2017). Extensions considered in such preliminary stage included altering existing images through rotation, resampling, and transposition techniques, as well as generating synthetic observations obtained by perturbing available images or training a support model for image generation (Goodfellow *et al.,* 2020).

3.4 One-shot learning and Siamese neural networks

Preliminary analysis suggested that data augmentation, in our case, is subject to the trade-off between the instability of the training process due to limited volumes of observations, and potential bias associated with synthetic data augmentation. In particular, data augmentation tends to expose the neural network to two types of issues. On one hand, additional images that over-represent the occurrence of printing defects with certain characteristics make the model exposed to overfitting those materialization of the defects only: this would undermine the model's ability to detect bite defects exhibiting different characteristics from those observed for training. On the other hand, the generation of novel synthetic defects could introduce relevant forms of algorithmic bias, in case these were unobservable in reality but relevant to steer the model's reasoning.

To curb such limitations, we resorted to CNN architectures following the *one-shot learning* approach (Fei *et al.,* 2006).

One-shot learning represents a specialization of statistical learning, aimed to emulate the ability of the human mind to associate entities unknown to the subject with known ones, based on similarity criteria. The approach thus aims to reproduce the ability of natural intelligence to make visual comparisons based on the knowledge of a ground truth conferred by experience. Given an image of a

generic entity - such as an object, a human face, or a landscape - and a range of possible categories of membership, the assignment of the image to one of these categories occurs from the simultaneous identification of points of commonality and points of difference with representative images of each class.

In its most rigorous formulation, one-shot learning requires that the training set consists of a number of prototypical observations equal to the number of categories of interest. However, the literature generally refers to one-shot learning in its more extended meaning[7] referred to as *k-shot*, in which a (limited) number of examples per class are considered, then differentiating it from *zero-shot* learning (Palatucci *et al.,* 2009) whereby no example image per class is available during training[8].

For one-shot learning, we make use of so-called *Siamese* neural networks (Koch *et al.,* 2015). Siamese neural networks exhibit two parallel sub-architectures (i.e., branches) identical to each other in terms of structure and parameter values (such as in green in Figure 1). Each branch of a Siamese network processes the respective element of the pair of observations that constitutes its input: a "verification" observation, that needs potential attribution to a given class, and a "support" observation, representing the class. The network maps the pair of observations toward "signals", synthetic and separable representations of the input. The intuition is as follows: two observations belonging to the same category are projected by the parallel architectures onto signals that are similar to each other; conversely, observations belonging to different categories will be characterized by orthogonal signals. The verification and support signals, respectively. $\boldsymbol{y}^{Sig,v}, \boldsymbol{y}^{Sig,s} \in \mathbb{R}^m$, are combined and processed by a "comparison" layer (Figure 1, in orange), which quantifies their similarity by the representation $\boldsymbol{y}^{L'} \in \mathbb{R}^{n_{L'}}$, obtained from the transformation:

$$\boldsymbol{y}^{L'} = g^{Sim}\left(\boldsymbol{w}^{Sim} Sim\left(\boldsymbol{y}^{Sig,v}, \boldsymbol{y}^{Sig,s}\right) + \boldsymbol{b}^{Sim}\right)$$

where g^{Sim} and *Sim* are an activation function and a similarity function, respectively, $\boldsymbol{w}^{Sim} \in \mathbb{R}^{n_{L'} \times m}$ is the matrix of weights and $\boldsymbol{b}^{Sim} \in \mathbb{R}^{n_{L'}}$ is the vector of bias terms. In the general case $L' \leq (L+1)$, that is an additional sequence of representation layers could be attached from the comparison layer. In the simplest case, considered by our application, $L' = L + 1$, i.e., the representation obtained from the output of the comparison layer is definitely the output of the Siamese network, such that, in our binary case study, $n_{L'} = 1$ and g^{Sim} is the sigmoid function that projects $\boldsymbol{y}^{L+1} = \boldsymbol{y}^{L'}$ in the interval [0,1]. In the training phase, let N be a pairs of images in the training set; each pair is labeled as containing images - one as verification and one as support - belonging to the same class ($c_i = 1$) or to distinct classes ($c_i = 0$). The values

[7] *One-shot* learning should be distinguished from "*one-shot* transfer" which has to be intended as a specialization of *transfer learning*.

[8] In the literature, there are numerous examples of the application of neural networks based on *zero-shot* learning, as well as in the area of natural language processing. Among the most frequent applications are classification exercises under conditions of extreme class imbalance (Ochal *et al.,* 2021), as well as face recognition tasks, which are required to guarantee a certain level of performance under conditions of high variability, for example in facial expression, brightness or background.

$y_1^{L+1}, \dots, y_N^{L+1}$ produced by the Siamese network therefore correspond to the probabilities that the *N* verification images belong to the same class as their respective support images. Training the model thus consists of minimizing the following *loss function*:

$$FP(\boldsymbol{y}^{L+1}) = -\frac{1}{N}\sum_{i=1,\dots,N} c_i \log\left(y_i^{L+1}\right) + (1-c_i)\log\left(1-y_i^{L+1}\right)$$

where c_i takes as said value 1 when the verification and supporting observations belong to the same class, 0 otherwise, and $y_i^{L+1} = Prob(c_i = 1), i = 1, \dots, N$.

4. Siamese networks for bite-type defect detection

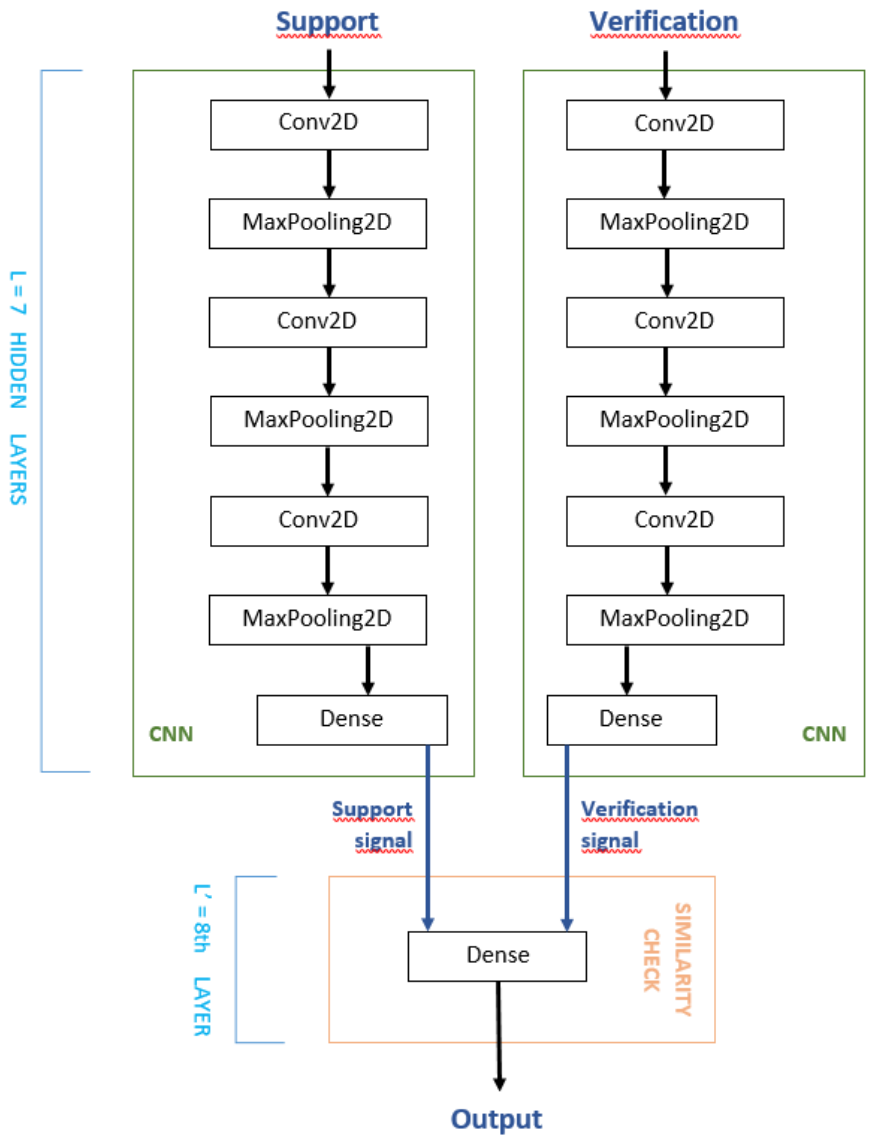
4.1 Model architecture and training

In the developing of a Siamese network for the detection of bite defects on the European flag graphical element, we resorted to an essential architecture consisting of a pair of twin networks as depicted in Figure 1. Each network is characterized by three convolutional layers (*Conv2D*) followed by three collection layers (*MaxPooling2D*), on which a sequence of fully connected layers (*Dense*) is grafted for signal vectors generation. These are combined according to the absolute distance metric and used to derive the output value $\boldsymbol{y}^{L+1}$, corresponding to the probability that the pair belongs to the same category[9].

[9] Additional variants were considered during the experimental phase. For brevity and little contribution in terms of relevance to the final considerations, the results are not reported.

Conceptual representation of the Siamese network architecture used in this work.

Figure 1



To facilitate comparison between results, the architecture of the networks is unchanged in the main blocks, while the number of processing units, i.e. nodes of each layer, depends on the resolution considered. Specifically, after of the manual acquisition of the banknotes high-resolution images (~2656x1467 pixels, 24-bit color, RGB model), the following four sets of images were generated by re-sampling, each of which was used in separate sessions for training and testing of the Siamese network in order to compare the performance obtained in the four cases:

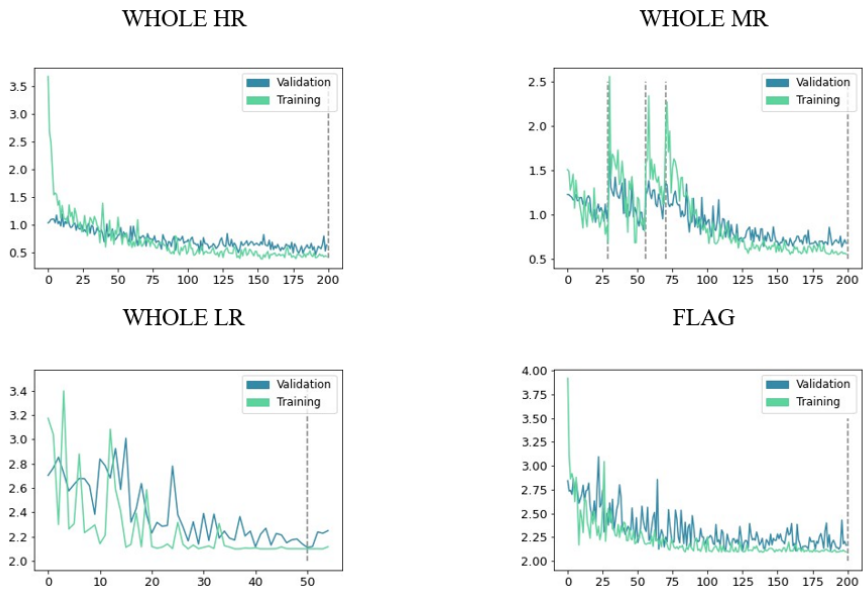
- Whole HR: whole banknote, high resolution (896x528 pixels);
- Whole MR: whole banknote, medium resolution (448x264 pixels);
- Whole LR: whole banknote, low resolution (299x176 pixels);
- Flag: cropping of the banknote's flag[10] in high resolution (224x352 pixels).

The training of the Siamese networks is done as mentioned through the Stochastic Gradient Descent (SGD) algorithm, starting from a set of parameters initialized according to the Xavier method. In order to promote optimal convergence of the process, batch normalization and dropout are also used, as well as re-initialization of the value of the parameters when a local optimum is reached, or the training set is overfitted (Figure 2). Hyper-parameters such as *Leaky ReLu* activation functions (Maas *et al.,* 2013) are also defined using a set of calibration images excluded from the training process.

[10] See Jang *et al.* (2020).

Siamese neural network training process representation for the four different input resolution considered: Whole HR, Whole MR, Whole LR and Flag. The green (blue) curves represent the value of the loss function during the iterations of the SGD optimization routine in the training (validation) set images. The dashed lines indicate the re-initialization of the training process, corresponding to a local optimum convergence.

Figure 2



The high performance of machine learning algorithms often has to contend with reduced interpretability of the learned internal mechanisms, as well as of the explainability of the underlying logic. Useful techniques to support the understanding, and thus the control of the proper logic of a complex analytical model, pertain to the field of so-called *eXplainable Artificial Intelligence* (Molnar, 2020). In the case of Siamese networks, the graphical representation of signals allows visual comparison between the synthesis of the supporting and verification images for a qualitative assessment of the similarity of the vectors extracted from each twin network. However, the logic underlying the signal extraction process is concealed by the large number of parameters used for recursive projection of the information content of the input images, consistent with the poor interpretability of deep learning models (Zhang *et al.,* 2019).

Unlike other types of neural networks, in CNNs the transmission of input images from one convolutional layer to another allows intermediate tensors to be extracted, which can in turn be represented as images, called *attention masks*. In the literature, the use of attention masks is typically used during model training, and is aimed at exogenous control of the process by the analyst (Xu *et al.,* 2015). Intuitively, the weight tensor of the l-th convolutional layer $\boldsymbol{w}^l$ is combined

element by element with an arbitrary mask $\boldsymbol{a}$ to emphasize or not emphasize specific areas of the image, and transformed as seen above according to the activation function g^l. The resulting tensor, $\boldsymbol{w}^{l,a}$, is thus characterized by elements such that higher values will correspond to greater network attention to a particular area of the image and vice versa.

For the purposes of our application, attention mechanisms are divided into *hard* and *soft*. Hard attention actually corresponds to random or deterministic exclusion of part of the surface (*cropping*). In this case $\boldsymbol{a}$ takes binary values: 0 or 1. Hard attention mechanisms can be used to exogenously direct the learning process or to maximize the exploration of the tensor surface and define optimal attention trajectories by the model (Ba *et al.,* 2014). In its simplest form, soft attention instead resorts to an activation function that compresses the values of the weights into a given range (Xu *et al.,* 2015), such as in the range [0,1] in the case of the *softmax* function. Built-in soft attention mechanisms are used to ensure overall control of the learning process[11] (Gregor *et al,* 2015; Kosiorek *et al.,* 2017).

This application considers an alternative use of soft attention masks according to a simplified approach geared toward model explicability. Attention masks are locally activated, after the training, to extract partial images along the twin CNNs, representing the logic adopted by the model along the processing steps. Specifically, for each convolutional layer the corresponding weight tensor is projected into the unit interval [0,1] according to the softmax activation function, which highlights the relative weight of each element. The weights thus transformed are then combined with the input of the convolutional layer. This corresponds to superimposing a soft attention mask on the banknote image. The intuition is as follows: the l-th layer uses weights whose relative values in the softmask are close to 1 at regions considered relevant for signal extraction, and close to 0 vice versa (see Figure 4).

4.2 Results

Having completed the learning phase, the Siamese network classifies each new image using the support image sets whose membership in one of the two categories of bite and no-bite is known by construction. Assignment of a new image to one of the two classes is then made from a comparison of the image and the two support sets.

The selection of the support images to be used for comparison is a critical step in the use of Siamese networks because classification is done from pairs of observations. The choice on the method for the support image selection certainly has an impact on the overall performance of the network; to analyze this impact, a hold-out image set, i.e., a portion of the images in the verification set, was selected, and initially separated from the training set, and finally used for the sole

[11] Specifically, at a given output, such as the classification of an element on the image surface, the model is asked to identify the reference region by operating a bias in the value of the corresponding parameters.

evaluation of the support image selection method to ensure the quality of the analysis. The following approaches are then considered:

1. RND-1: Each verification image is compared with one randomly selected support image per class;

2. SSIM-0: Each verification image is compared with a specific single support image per class: the support selected image is the one with the maximum similarity with the verification image among those available in that particular class, according to the Structure Similarity Index Measure (SSIM)[12];

3. SSIM-1: Each verification image is compared with the second most similar of the available support images candidates for each class. In the SSIM-1 approach, the exclusion of the most similar support image is intended to reduce the impact of not pertinent but similar features that might be misleading for the classification exercise;

4. SSIM-k: Each verification image is compared with k suppor images per class, selected according to similarity (k=10 in the following analysis).

Table 1 shows the results obtained using the four different resolutions described above, and using the four different selection methods for the support images. The test set contains 46 images, including 14 with bite defect of different severity[13]. The performance on the test set is measured from the Recall rate[14], i.e., the rate of images with correctly classified bite defect, and the F1-score, the harmonic mean of the *Recall* and *Precision* score, defined as the rate of banknotes correctly classified as not defective. In our application, false positives, i.e., images of banknotes incorrectly identified as defective, and defect under-reporting, or false negatives, are considered equally costly.

[12] SSIM is a widely used measure for image comparison and image quality assessment, introduced by Wang *et al.* (2004). It is derived from the composition of three comparison runs related to image brightness, contrast and texture.

[13] The classification is based on the documentation provided by the European Central Bank.

[14] Given the number of banknotes classified as defective by the network (*all positives*), *Recall* is defined by the proportion of banknotes correctly classified (*true positives*): *Recall = (true positives)/(all positives)*.

Results for the binary image classification exercise in the test set, for the high (HR), medium (MR) and low (LR) resolution whole banknote cases, and for the European flag crop case. Performance below the threshold value 0.5 is indicated by "-".

Table 1

	Recall				F1-Score			
	RND-1	**SSIM-0**	**SSIM-1**	**SSIM-k**	**SSIM-k**	**SSIM-0**	**SSIM-1**	**SSIM-k**
Whole HR	-	.75	.62	-	-	.67	.55	-
Whole MR	1.	**.87**	**.87**	.5	.84	.82	**.93**	.67
Whole LR	-	.62	.62	-	-	.55	.62	-
Flag	-	-	.5	.98	-	-	**-**	-

We observe that, for the verification set considered, the use of the medium-resolution whole image (Whole MR) allows for greater accuracy in classifying observations into bite and no-bite. In particular, the SSIM-1 selection method appears to perform better in assigning the banknote to the correct category.

Siamese networks are characterized by their ability to identify critical features during the projection of the input into the signal, thus the assignment of the verification image to its correct class, under conditions of high variability. When this variability is quite excessive, as in the case of high resolution images (Whole HR) the granularity in terms of pixels becomes misleading to separate signals obtained from the verification and the support images. This can also be seen from the visual inspection of the signals extracted from the Siamese network, shown in the top left panel of Figure 3, where it can be seen that signals are more similar to each other when compared with the medium-resolution case (top right). Similarly, by excessively reducing the image resolution, the absolute size of the bite defect, the relative size of the field corresponding to the flag, and the overall variability turn out to be attenuated to a point that the Siamese network cannot learn a reliable criterion for their recognition (bottom left). In the case of cropped images containing the flag element only, the combination of the reduction in overall (second-level) variability and the increased relative weight of the region of pixels corresponding to the defect, which is characterized by high (first-level) variability, make a one-shot learning approach inappropriate. Indeed, the

performance measured in quantitative terms, and the graphical inspection of the obtained signals (bottom right) reflect this inadequacy.

Comparison of signals for verification and support images. The images show the signals obtained for pairs of images of type Whole HR, Whole MR, Whole LR, and Flag. Support images were selected using the SSIM-1 method. For each box, the verification signal is obtained from an image with a bite defect, to be compared with the bite and no-bite type support images, respectively. Figure 3

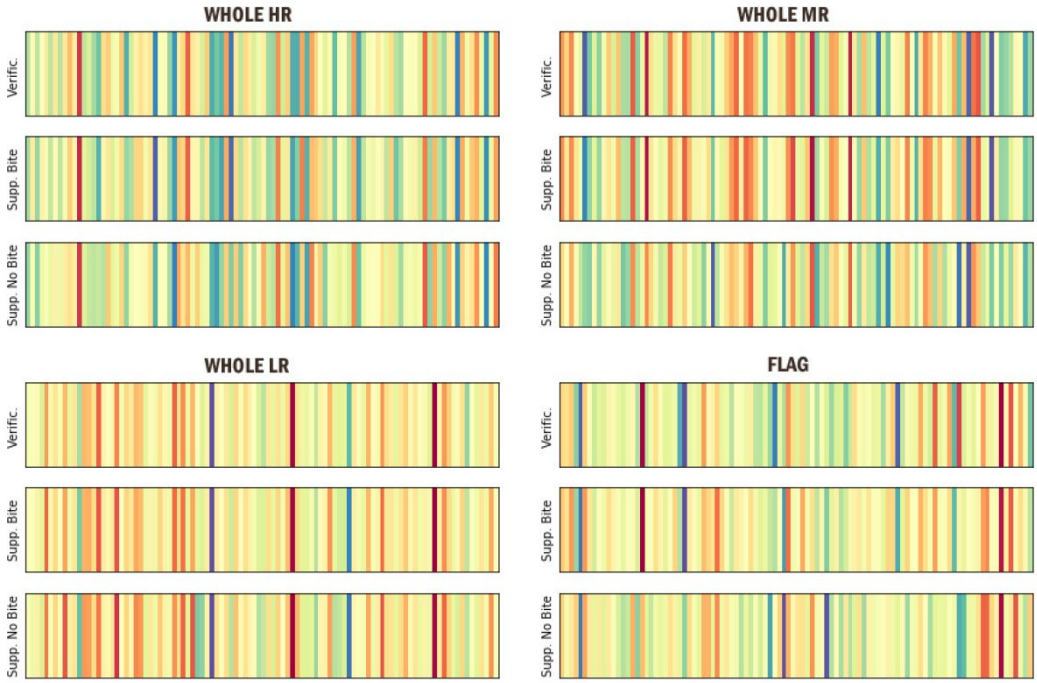


Figure 4 illustrates an example of the attention mechanism operated by the convolutional components of a Siamese networks, obtained for the same bite defect image, and in the four cases Whole HR, MR, LR, and Flag crop. In the different images, brighter shades correspond to a greater attention, or concentration, conferred by the neural network in the three convolutional layers inside the CNN. It can be seen that, in the absence of indications given to the network during its construction, both the architecture for the analysis of high-resolution images and that for medium-resolution images are able to independently identify the area of the critical surface for our purposes of bite defect detection, and subsequent assignment to one class or the other. This virtuous process does not seem to be triggered, except approximately, in the case of the Siamese network calibrated from the low-resolution training set. Similarly, the same behaviour is produced in the cropped flag case: the Siamese network

seems to "focus" on the entire image area, thus making the process of synthesizing the input information overly susceptible to minor variations.

Graphical representation of the attention mechanism operated by the three blocks of the convolutional component of the Siamese networks for a training image with bite defect, in the formats Whole HR, Whole MR, Whole LR, and Flag. Figure 4

5. Conclusions and possible extensions

This paper reports the main results obtained from the exploratory study on the use of neural networks for the autonomous recognition and classification of banknote defects in the context of quality control. The complexity of the problem, mainly due to the high degree of natural product variability and the variability of defect types and locations, imposed a very stringent approach in the selection of experimental scenarios. We restricted the analysis to a specific type of defects

(bites), located on a predetermined portion (Flag) of a single denomination of banknotes (€50 of the second series).

The results support the relevance of these tools for identifying print defects. In particular, the case of medium-resolution whole banknote (Whole MR) allows balancing the needs related to image acquisition, computational efficiency, and quality of model performance. Further work needs to be carried out in order to attest their actual potential as an alternative, or cooperative, use along with current non-automated qualitative techniques. Further study and experimentation activities must extend the analysis to categories of defects other than bite, making use of useful techniques to adapt a previously trained model to a new task (so-called *transfer learning*), while maintaining a binary classification output. Finally, the implementation of a system for the classification of multiple print defects could make use of so-called *ensemble learning* techniques to combine and aggregate the results produced by a multi-network architecture.

However, any extensions considered would be circumscribed by an a-priori choice that keeps their domain of analysis and research within one-shot learning algorithms, and specifically using Siamese networks. This choice is also suggested by the literature produced by the scientific community for problems variously related to what is presented in these pages.

References

Ali, T., Jan, S., Alkhodre, A., Nauman, M., Amin, M., & Siddiqui, M.S. (2019). DeepMoney: counterfeit money detection using generative adversarial networks. *PeerJ Computer Science*, *5*, e216.

Aseffa, D.T., Kalla, H., & Mishra, S. (2022). Ethiopian banknote recognition using convolutional neural network and its prototype development using embedded platform. *Journal of Sensors*, *2022*, 1-18.

Ba, J.L., Mnih, V., & Kavukcuoglu, K. (2014). Multiple object recognition with visual attention.

Boureau, Y.L., Ponce, J., & LeCun, Y. (2010). A theoretical analysis of feature pooling in visual recognition. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 111-118).

Darade, S.R., & Gidveer, G.R. (2016). Automatic recognition of fake Indian currency note. In *2016 international conference on Electrical Power and Energy Systems (ICEPES)* (pp. 290-294). IEEE.

Daraee, F., & Mozaffari, S. (2010). Eroded money notes recognition using wavelet transform. In *2010 6th Iranian Conference on Machine Vision and Image Processing* (pp. 1-5). IEEE.

Desai, S., Rajadhyaksha, A., Shetty, A., & Gharat, S. (2021). CNN based counterfeit Indian currency recognition using generative adversarial network. In *2021*

International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 626-631). IEEE.

Domínguez, F.R. (2020). Supermasks and a Good Initialization Are All You Need (*Doctoral dissertation, Pontificia Universidad Catolica de Chile (Chile)*).

Fei-Fei, Li, Rob Fergus e Pietro Perona (2006). One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence 28.4*: 594-611.

Feng, B.Y., Ren, M., Zhang, X.Y., & Suen, C.Y. (2014). Part-based high accuracy recognition of serial numbers in bank notes. In *Artificial Neural Networks in Pattern Recognition: 6th IAPR TC 3 International Workshop, ANNPR 2014, Montreal, QC, Canada, October 6-8, 2014. Proceedings 6* (pp. 204-215). Springer International Publishing.

García-Lamont, F., Cervantes, J., López, A., & Rodríguez, L. (2013). Classification of Mexican paper currency denomination by extracting their discriminative colors. In *Advances in Soft Computing and Its Applications: 12th Mexican International Conference on Artificial Intelligence, MICAI 2013, Mexico City, Mexico, November 24-30, 2013, Proceedings, Part II 12* (pp. 403-412). Springer Berlin Heidelberg.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249-256). JMLR Workshop and Conference Proceedings.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. *MIT press*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139-144.

Gregor, K., Danihelka, I., Graves, A., Rezende, D., & Wierstra, D. (2015). Draw: A recurrent neural network for image generation. In *International conference on machine learning* (pp. 1462-1471). PMLR.

Grijalva, F., Rodriguez, J.C., Larco, J., & Orozco, L. (2010). Smartphone recognition of the US banknotes' denomination, for visually impaired people. In *2010 IEEE ANDESCON* (pp. 1-6). IEEE.

Guo, Z.H., & Li, S.H. (2006). Feed-Forward neural network using SARPROP algorithm and its application in radar target recognition. In *Advances in Neural Networks-ISNN 2006: Third International Symposium on Neural Networks, Chengdu, China, May 28-June 1, 2006, Proceedings, Part II 3* (pp. 369-374). Springer Berlin Heidelberg.

Hasanuzzaman, F.M., Yang, X., & Tian, Y. (2011). Robust and effective component-based banknote recognition by SURF features. In *2011 20th Annual Wireless and Optical Communications Conference (WOCC)* (pp. 1-6). IEEE.

Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R.R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448-456). pmlr.

Jadhav, M., kumar Sharma, Y., & Bhandari, G.M. (2019). Currency identification and forged banknote detection using deep learning. In *2019 International Conference on Innovative Trends and Advances in Engineering and Technology (ICITAET)* (pp. 178-183). IEEE.

Jang, U., Suh, K.H., & Lee, E.C. (2020). Low-quality banknote serial number recognition based on deep neural network. *Journal of Information Processing Systems*, *16*(1), 224-237.

Kamble, K., Bhansali, A., Satalgaonkar, P., & Alagundgi, S. (2019). Counterfeit currency detection using deep convolutional neural network. In *2019 IEEE Pune Section International Conference (PuneCon)* (pp. 1-4). IEEE.

Ke, W., Huiqin, W., Yue, S., Li, M., & Fengyan, Q. (2016). Banknote image defect recognition method based on convolution neural network. *International Journal of Security and Its Applications*, *10*(6), 269-280.

Khemiri, W., Tarifa, A., Jaafar, W., & Abderrazak, J.B. (2022). Towards a Hybrid Variant of GAN For Counterfeit Money Detection SSQUGAN: A Semi-Supervised Quadrupled GAN.

Koch, G., Zemel, R., & Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop* (Vol. 2, No. 1).

Kosiorek, A., Bewley, A., & Posner, I. (2017). Hierarchical attentive recurrent tracking. *Advances in neural information processing systems*, *30*.

Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90.

Laavanya, M., & Vijayaraghavan, V. (2019). Real time fake currency note detection using deep learning. *Int. J. Eng. Adv. Technol.(IJEAT)*, *9*.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, *2*.

Lee, J. W., Hong, H.G., Kim, K.W., & Park, K.R. (2017). A survey on banknote recognition methods by various sensors. *Sensors*, *17*(2), 313.

Linkon, A.H.M., Labib, M.M., Bappy, F.H., Sarker, S., Jannat, M.E., & Islam, M.S. (2020). Deep Learning Approach Combining Lightweight CNN Architecture with Transfer Learning: An Automatic Approach for the Detection and Recognition of Bangladeshi Banknotes. In *2020 11th International Conference on Electrical and Computer Engineering (ICECE)* (pp. 214-217). IEEE.

Liu, L., Ye, Y.T., Xie, Y., & Pu, L. (2010). Serial number extracting and recognizing applied in paper currency sorting system based on RBF Network. In *2010 international conference on computational intelligence and software engineering* (pp. 1-4). IEEE.

Maas, A.L., Hannun, A.Y., & Ng, A.Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml* (Vol. 30, No. 1, p. 3).

Manikandan, K., & Sumithra, T. (2015). Currency recognition in mobile application for visually challenged. *Discovery*, *30*, 245-248.

Molnar, C. (2020). Interpretable machine learning. *Lulu.com*.

Mousavi, S.A., Meghdadi, M., Hanifeloo, Z., Sumari, P., & Arshad, M.R.M. (2015). Old and worn banknote detection using sparse representation and neural networks. *Indian J. Sci. Technol*, *8*, 913-918.

Ochal, M., Patacchiola, M., Storkey, A., Vazquez, J., & Wang, S. (2021). Few-shot learning with class imbalance. *arXiv preprint arXiv:2101.02523*.

Palatucci, M., Pomerleau, D., Hinton, G.E., & Mitchell, T.M. (2009). Zero-shot learning with semantic output codes. *Advances in neural information processing systems*, *22*.

Park, C., Cho, S.W., Baek, N.R., Choi, J., & Park, K.R. (2020). Deep feature-based three-stage detection of banknotes and coins for assisting visually impaired people. *IEEE Access*, *8*, 184598-184613.

Pham, T.D., Lee, D.E., & Park, K.R. (2017). Multi-national banknote classification based on visible-light line sensor and convolutional neural network. *Sensors*, *17*(7), 1595.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural networks*, *12*(1), 145-151.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

Sawant, V.M., Tupe, R. D., Tawade, A.S., & Bhalerao, S.M. (2022). Fake Currency Identification System Using CNN. *International Journal of Wireless Network Security*, *8*(1), 18-23.

Sharma, B., & Kaur, A. (2012). Recognition of Indian paper currency based on LBP. *International Journal of Computer Applications*, *59*(1).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sun, B., & Li, J. (2008). The recognition of new and old banknotes based on SVM. In *2008 Second International Symposium on Intelligent Information Technology Application* (Vol. 2, pp. 95-98). IEEE.

Suresh, I.A., & Narwade, P.P. (2016). Indian currency recognition and verification using image processing. *International Research Journal of Engineering and Technology (IRJET)*, *3*(6), 87-91.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

Treadgold, N.K., & Gedeon, T.D. (1996). The SARPROP algorithm, a simulated annealing enhancement to resilient back propagation. In *Proceedings International Panel Conference on Soft and Intelligent Computing* (pp. 293-298).

Veeramsetty, V., Singal, G., & Badal, T. (2020). Coinnet: platform independent application to recognize Indian currency notes using deep learning techniques. *Multimedia Tools and Applications*, *79*(31-32), 22569-22594.

Wang, J., & Perez, L. (2017). The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks Vis. Recognit*, *11*, 1-8.

Wenhong, L., Wenjuan, T., Xiyan, C., & Zhen, G. (2010). Application of support vector machine (SVM) on serial number identification of RMB. In *2010 8th World Congress on Intelligent Control and Automation* (pp. 6262-6266). IEEE.

Wijnhoven, R.G., & de With, P.H.N. (2010). Fast training of object detection using stochastic gradient descent. In *2010 20th International conference on pattern recognition* (pp. 424-427). IEEE.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R. & Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning* (pp. 2048-2057). PMLR.

Yan, Z., Zhang, H., Piramuthu, R., Jagadeesh, V., DeCoste, D., Di, W., & Yu, Y. (2015). HD-CNN: hierarchical deep convolutional neural networks for large scale visual

recognition. In *Proceedings of the IEEE international conference on computer vision* (pp. 2740-2748).

Yu, K., Xu, W., & Gong, Y. (2008). Deep learning with kernel regularization for visual recognition. *Advances in neural information processing systems*, *21*.

Wang Z., Bovik, A.C., Sheikh H.R., Simoncelli, E.P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing 13.4*: 600-612.

Zeiler, M.D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13* (pp. 818-833). Springer International Publishing.

Zhang, Q., Yan, W.Q., & Kankanhalli, M. (2019). Overview of currency recognition using deep learning. *Journal of Banking and Financial Technology*, *3*, 59-69.

Siamese neural networks for detecting banknote printing defects

K. Boria, A. Luciani, S. *Marchetti* and M. Viticoli
(Bank of Italy)

3rd IFC Workshop on Data Science in Central Banking
Oct. 18, 2023

Introduction

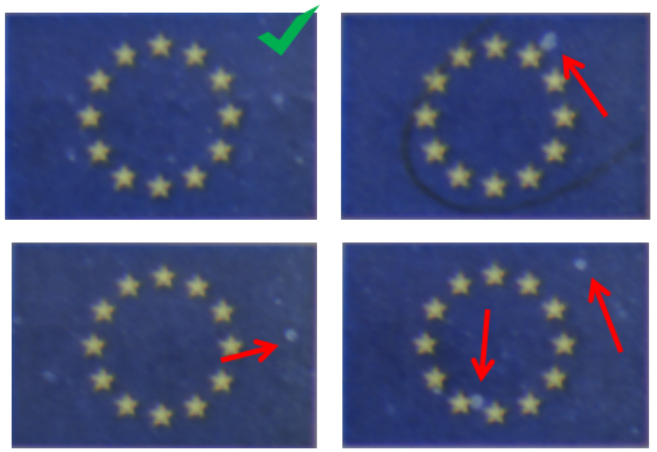
- Banknote production in the Eurosystem relies on strict quality controls throughout the process (ECB quality requirements, ISO standards);
- Similar to any manufacturing process, the printing of banknotes can give rise to various imperfections and defects;
- The number, type, and size of defects on banknotes are critical for the conformity of production batches (with related losses).

Introduction

- Banknotes exhibit variability in the position of elements, and in the shade and intensity of colors;
- While measurement of many parameters is carried out via automatic optical systems, the validation process cannot be fully automatized and requires highly trained staff (potential subjectivity factors);
- Artificial Intelligence (AI) systems are profitably used across industries to support and automatize quality control;
- Our work focuses on a denomination/defect pair to assess whether neural networks could improve process efficiency.

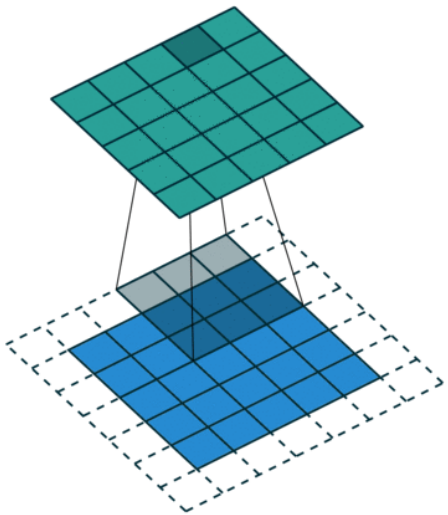
Data

- Since acquisition of high-resolution images is costly:
  - Focus on €50 banknotes of the *Europa* series;
  - Focus on flag-*bite* defects: lack of ink on a homogenous background, with varying shape, size, and position.
- We manually acquire 24-bit color images (RGB model) of banknotes with 2,656 x 1,467 pixels resolution and annotate them as *fit*/*unfit*.



Neural Networks & Image Processing

- Previous research on banknotes analysis mostly resorts to convolutional neural networks (CNNs);
- Applications address recognition of denomination, serial number, currency, state of wear as well as counterfeit detection. On the production phase, we mention Pham et al., 2017; Ke et al., 2016.
- The inherent variability of banknotes requires training of complex CNN architectures, usually addressed via data augmentation.

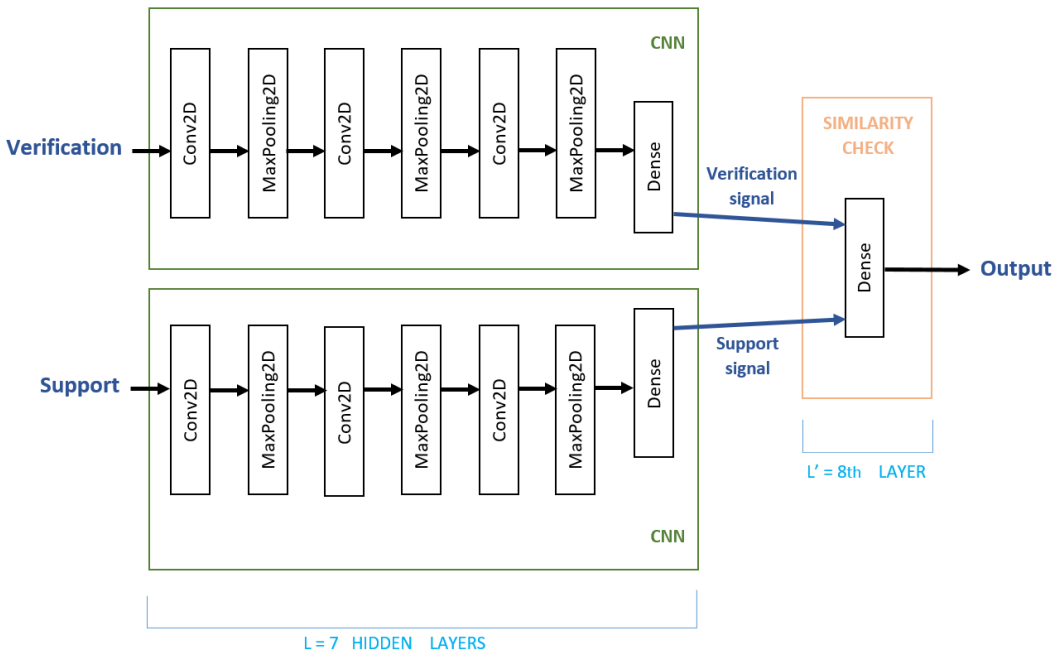


One-shot learning & Siamese Networks

- The traditional way: data augmentation
  - Manual: Overfitting of specific bite defects;
  - AI-enabled: Hallucination-prone behavior;
- The practical way: *one-shot learning* (Fei et al., 2006)
  - Emulates the ability of the human mind to associate entities based on similarity criteria.

Model development & Training

- We build a Siamese neural network architecture (Koch et al., 2015) for the bite-detection task:
  - Each twin branch extracts a signal from its input, either a *verification* or a *support* image;
  - The more similar the signals, the most likely images belong to the same class.

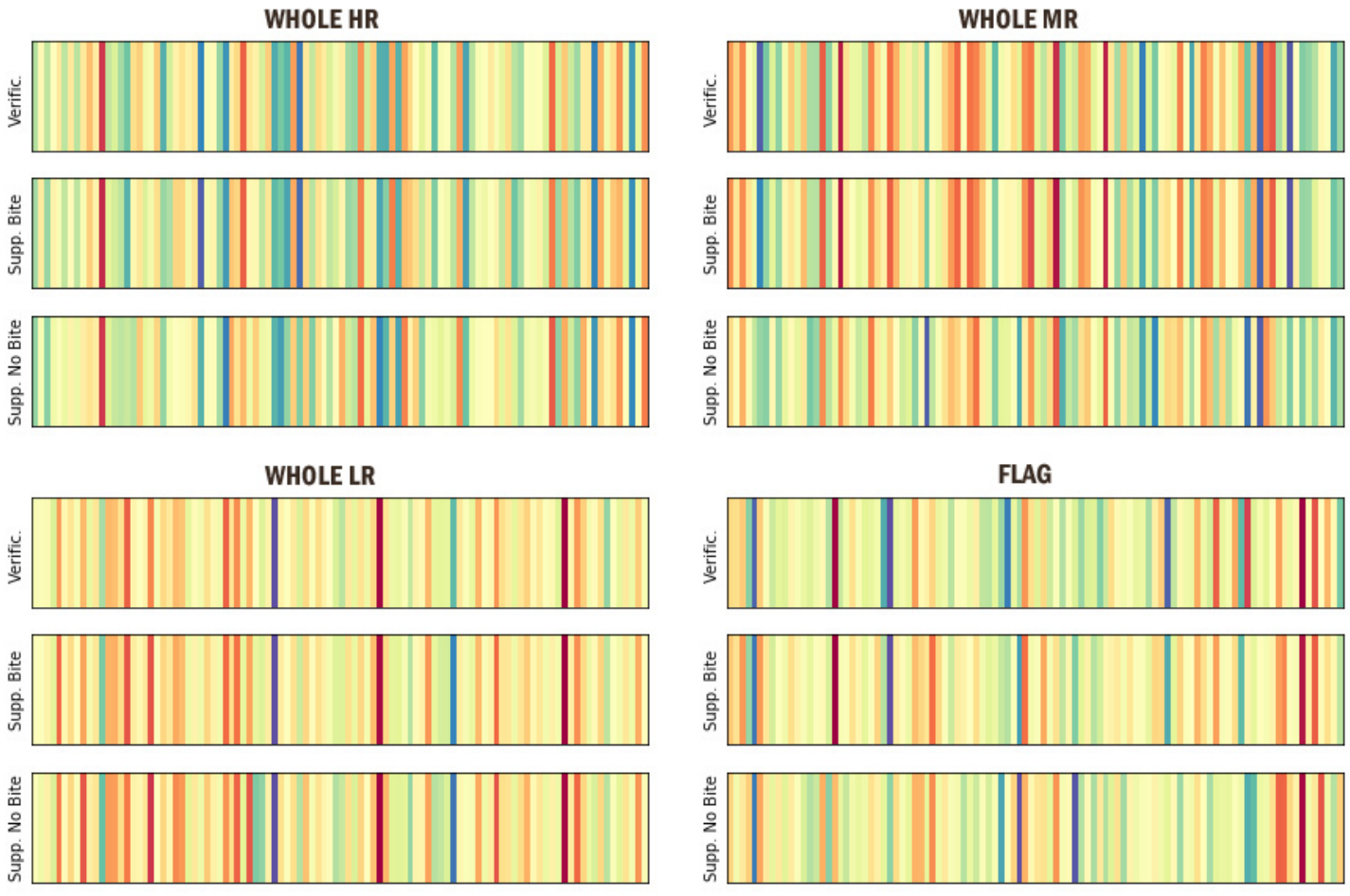


Model development & Training

We train our models on different resolutions of the input set: high, medium and low resolution + cropped flag detail:

- Training set: $N = 200$ images (100 bites, 100 fit; $C(N, 2)$ training pairs);
- We ensure convergence of the training process to global minima via drop-out, batch normalization, random re-initialization of weights;
- We control internal logics of networks via soft masks extracted from convolutional layers.

Empirical Results

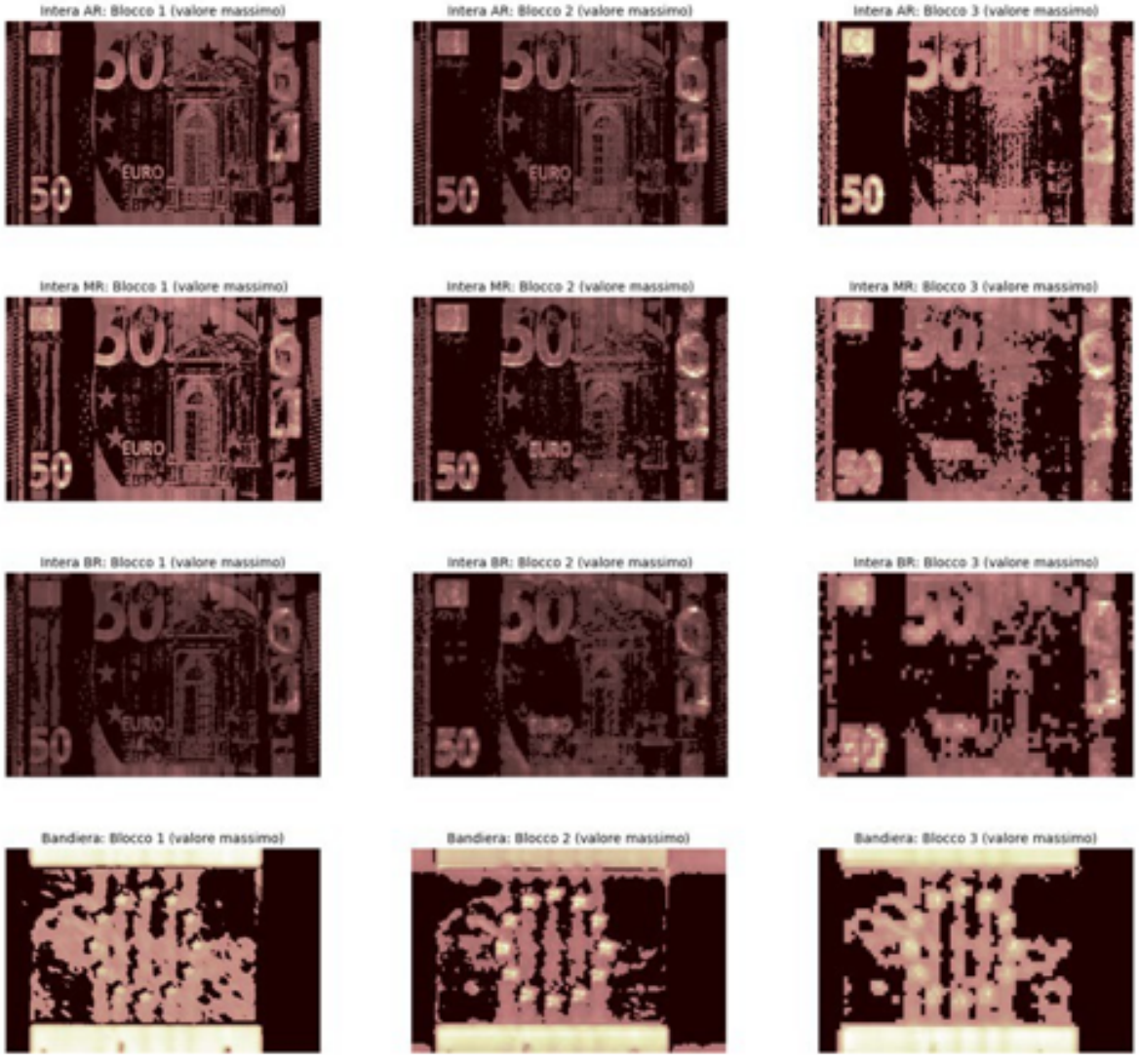


Empirical Results

	Recall				F1-Score			
	RND-1	SSIM-0	SSIM-1	SSIM-k	SSIM-k	SSIM-0	SSIM-1	SSIM-k
WHOLE HR	-	.75	.62	-	-	.67	.55	-
WHOLE MR	1.	.87	.87	.5	.84	.82	.93	.67
WHOLE LR	-	.62	.62	-	-	.55	.62	-
FLAG	-	-	.5	.98	-	-	-	-

Test set size: 46 images (14 bites; 32 fit);
RND-1: Support image is picked at random;
SSIM-j, j=0,1: the (j+1)th most similar image is picked as support;
SSIM-k: the k most similar images are picked as support, with k=10.

Empirical Results



Conclusions & Next Steps

- Our exploratory study provides insights on the detection of banknote defects for quality control via one-shot learning;
- We find medium resolution input allows for greater accuracy based on different metrics and evaluation criteria;
  - Trade-off between resolution and attention scattering due to variability patterns of banknotes;
- We are able to gain insights on the internal logics adopted by the models via soft masks;
- Future work will extend the analysis to additional defects (minor bite type defects already tested) and denominations, and to additional model architectures.

Thank you!