A neural classifier to evaluate the role of image resolution in the perception of color differences

Beatrice Lazzerini, and Francesco Pistolesi
Department of Information Engineering, University of Pisa
Largo L. Lazzarino, 1 – 56122 Pisa, ITALY
beatrice.lazzerini@unipi.it, francesco.pistolesi@unipi.it

***Abstract*—Measuring the color difference between an image and a copy is crucial in the color industry, as many printing processes exist whose aim is to accurately reproduce colors. Due to various phenomena that are often unpredictable, these processes may print copies that are perceptually different from the original. Visual inspections are thus required in order to constantly control the color quality. These inspections are made by experts who observe a master image and a copy, typically at high resolution. This takes a long time for image acquisition, and generates high costs that could be reduced if these processes could be performed at lower resolution, possibly without human intervention, as people may perceive colors in different ways. This paper presents a neural network that assesses the color difference of multicolored images (photos) at different resolutions. The results showed that the level of color difference perceived by the network remains unchanged as the resolution decreases, and sometimes is even more evident.**

***Index Terms*—Artificial intelligence, Classification, Color, Industry 4.0, Machine learning, Neural network, Perception.**

I. Introduction

The processes that involve printing are widespread in industries like packaging, commercial and decorative printing. The printing industry generates incredible revenues that achieved $980 billion in 2018 [1].

Printing processes require dedicated operators who constantly control the color fidelity because these processes are subject to several phenomena that impact on the rendering of colors. This may cause the production of batches—for example advertising posters, flyers, etc.—whose items are not chromatically homogeneous. Even worse, the printed colors may be different from those chosen by the customer when signing the supply contract. As a consequence, companies may be subject to huge financial losses: customers rarely purchase defective products. Controlling the color quality is thus crucial in the industry.

Nowadays, much of the color control in industry is still based on visual inspections performed by experts. However, people show different levels of color sensitiveness that depend on perceptual and cultural aspects, like mood, age, or, for example, the fact that warm/cool colors are observed [2]. Visual inspections may thus cause flaws that lead to reproducing non-compliant or undesired colors.

In the last years, the CIE (*Commission Internationale de l'Eclairage*, i.e., the international authority of light, illumination, color, and color spaces) has released the CIELAB color space, along with different indicators specifically aimed at objectively measuring the difference between two colors [3], [4]. These indicators depend on several parameters and weighting functions. Unfortunately, when dealing with particular colors such as blue tones, dark shades, grays and others, these indicators are not in compliance with the differences that the human eye sees [5].

Many techniques have recently been proposed to guarantee a high accuracy when reproducing colors. For example, a neural system was proposed in [6] to control the fidelity in solid-color prints (color patches), i.e., prints characterized by only one color with the same tone in all the image. Various studies on color calibration [7] have also been made, as well as others on the color appearance on electronic devices [8]. Other approaches investigated how the paper properties affect the color reproduction in digital printing [9]. Further approaches proposed functions to enhance the contrast and saturation [10]. There also exist several applications to guarantee that colors are reproduced accurately. Some interesting recent applications include spectrophotometric analysis of the color of ceramic restorations in dentistry [11], and pattern matching methods to control the quality of metallic coatings [12].

The evaluation of the color difference between two images is performed by operators, at high resolution. This entails purchasing and maintaining expensive equipment. Moreover, when working at high resolution, the image acquisition is slow, and needs large amounts of memory. Finally, as the evaluation is performed by one or more operators, the final result has high chances of being influenced by subjectivity.

This paper proposes a neural network that automatically measures the color difference between two multi-color images (photos). The neural network is used to measure the color difference at resolutions that are lower than that used in the printing process, in order to investigate how the perceived difference varies as the resolution decreases.

The paper is organized as follows. Section II contains a background on color and the basic concepts of image resolution. Section III presents the dataset. Section IV presents the method, i.e., the neural network and its architecture. Section V discusses the results. Section VI draws the conclusions.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

II. BACKGROUND

This section introduces some basic concepts of color representation, perception and resolution.

A. CIELAB

CIELAB is a color space released by the International Commission on Illumination (CIE) in 1976. It represents a color by using three values: $L*$ for the lightness from black (0) to white (100), $a*$ from green (–) to red (+), and $b*$ from blue (–) to yellow (+). The color space is shown in Fig. 1.

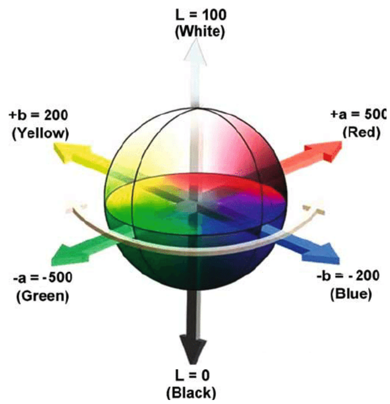

Fig. 1. The CIELAB color space.

CIELAB was purposely designed so that a given level of change of one or more coordinates among L*, a* and b* corresponds to a given level of visually-perceived difference. Unlike the RGB and CMYK color models, CIELAB was designed to reproduce the human vision, as it aspires to perceptual uniformity.

B. Image resolution

The resolution of an image is the number of pixels per inch, measured in *dpi* (dots per inch). Images with high dpi have more color information, which results in higher quality and larger files.

Several compression techniques have been developed over the years. However, the loss in information that occurs during compression may affect the color quality. For example, the JPEG algorithm is a good trade-off between quality and size (see Fig. 2) that uses two compression methods, one based on the discrete cosine transform with ”lossy” compression, and the other based on the use of a predictive method with lossless compression.

In this paper, the following 5 resolutions were considered:

- 4K $\rightarrow$ 3840x2160;
- 3K $\rightarrow$ 3000x2000;
- 1080p $\rightarrow$ 2048x1080;
- 720p $\rightarrow$ 1280x720;
- 480p $\rightarrow$ 544x480;

where each pair represents the width and height of the image. The product of these two values is the total number of pixels of the image. For example, at a resolution of 4K, the image has 3840*2160 = 8,294,400 pixels, whereas at 480p the number of pixels is 544*480 = 261,120.

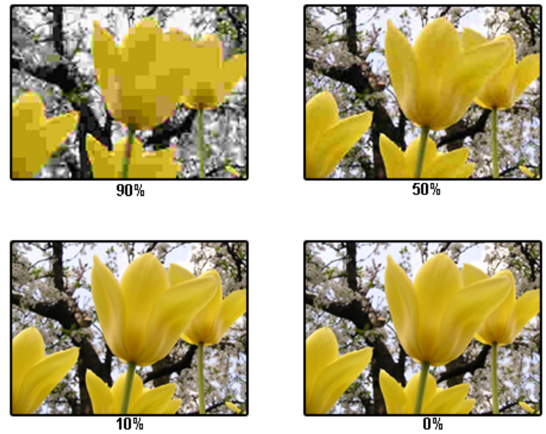

Fig. 2. An example of JPEG compression at different levels.

III. DATASET

The dataset used in the experiments was generated by an *initial dataset* made up of 40 master images in the CIELAB color space, at a resolution of 4K. These images are shown in Fig. 3.

Altered copies of the master images were generated according to a procedure with the following steps:

1) for each 4K image of the initial dataset, generate four *compressed copies* at resolutions 3K, 1080p, 720p and 480p;
2) apply three color filters with increasing intensities to each 4K image and its compressed copies.

The procedure is shown in Fig. 4.

Two of the filters applied at step 2) affect all the image, whereas the third filter only alters the regions with colors of a given tone. It was decided to alter the blue tones, as this color characterizes many images of the dataset.

A total of 2000 images were generated according to the previous procedure, which was implemented as a C++ algorithm based on the OpenCV library. Fig. 5 shows some examples of images generated with the previous procedure, by using filters at high intensity in order to make it more evident the effect of the filters.

IV. METHOD

This section describes the procedure to design the neural network that automatically measures the color difference between two images. The network was designed as a 5-class classifier that takes a set of features extracted from a master image and a copy, and returns the level of color similarity of the two images expressed as one of the five classes, namely, *Low*, *Medium-Low*, *Medium*, *Medium-High*, and *High*.

A. Feature extraction

Each image generated as explained in the previous section was divided into 9 contiguous squares of equal size. For each square, mean and variance were calculated for each of the 3

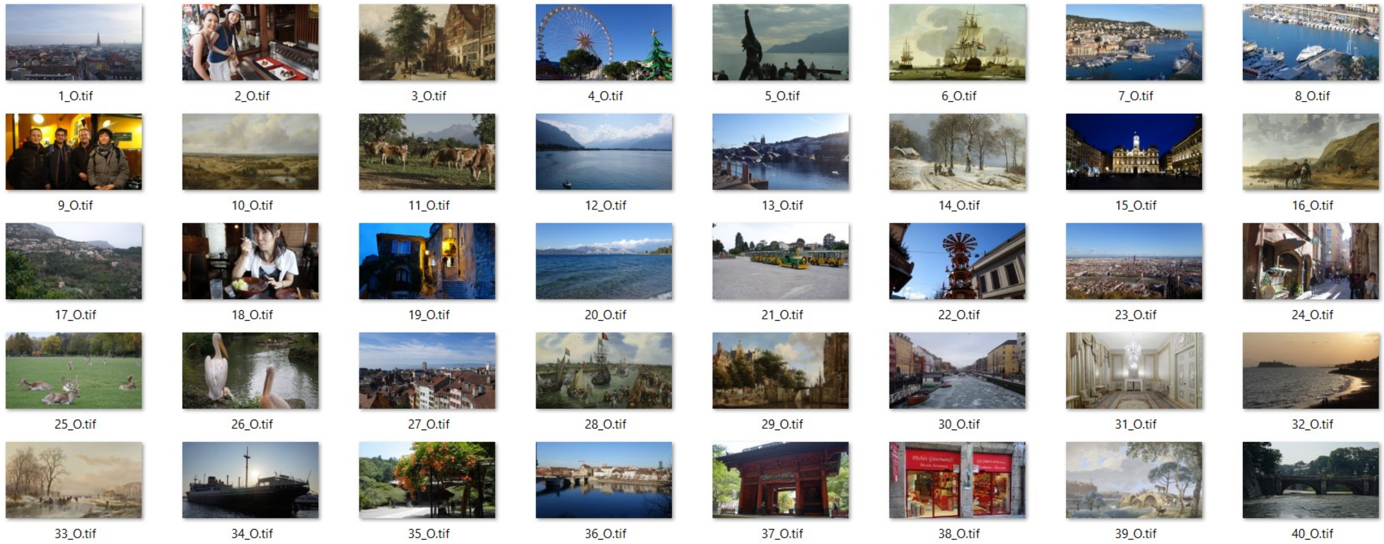

Fig. 3. Dataset used in the experiments.

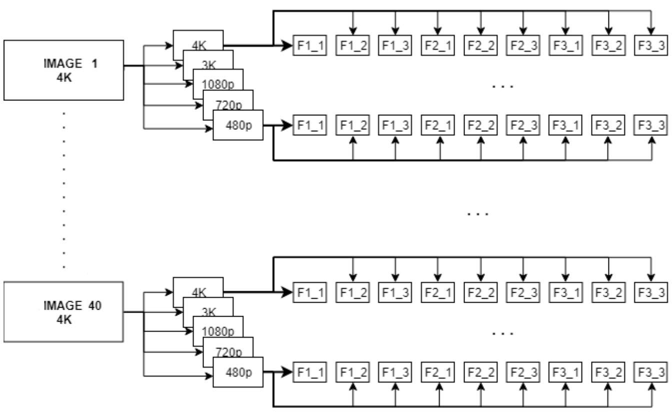

Fig. 4. Procedure to generate the dataset. Fj_α denotes a copy generated by applying filter j at intensity α.

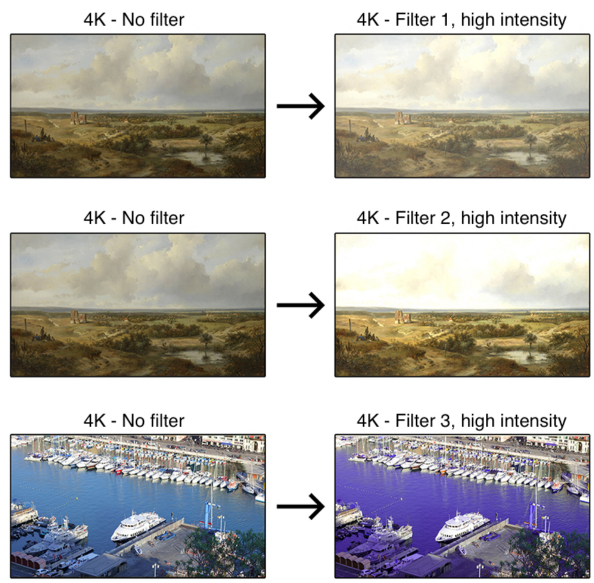

Fig. 5. Examples showing the effect of applying some of the filters to the images of the initial dataset.

channels (L*, a* and b*). The mean was calculated by using the `mean()` function of OpenCV, whereas the variance was calculated by implementing a dedicated function. Mean and variance were chosen as they are typically used in colorimetry for similar purposes, such as, for example, in the calculation of the C-SSIM index, which is a version of the SSIM index that compares the color and structure of two images [13].

The neural network thus takes as input the features of a master image, along with the features extracted from a copy. After calculating these values for all the master-copy pairs, the values were put into a 1800x108 size matrix, where 1800 is the number of pairs $(i, j(i, \alpha))$, where i denotes the image and $j(i, \alpha)$ denotes a copy of i generated by applying filter j with intensity α, whereas 108 is the total number of features of each master-copy pair.

In detail, given X and Y, respectively master and copy, and being $n = 9$ the number of squares into which images X and Y are divided, one row of the matrix contains, in the order:

- ${\mu_X}_k(C)$: average of each square of X in the 3 channels;
- ${\mu_Y}_k(C)$: average of each square of Y in the 3 channels;
- ${\sigma^2_X}_k(C)$: variance of each square of X in the 3 channels;
- ${\sigma^2_Y}_k(C)$: variance of each square of Y in the 3 channels;

with $k = 1, ..., n$, and channels $C \in \{$L*, a*, b*$\}$. The total number of features Φ is:

$$\Phi = 2 \times (3 \text{ channels} \times 9 \text{ squares} \times 2 \text{ values}) = 108. \quad (1)$$

Each row is thus a sample that represents the color information of a master-copy pair as a vector of Φ features.

B. Output

The output of the network is the level of color similarity of a master image compared to a copy. As introduced before, five increasing levels of color similarity were considered: Low (L), Medium-Low (ML), Medium (M), Medium-High (MH), High (H). The network is thus a five-class classifier.

A total of 1800 master-copy pairs of images were visually inspected (40 images $\times$ 9 copies $\times$ 5 resolutions) by ten observers who expressed a level of color similarity among Low (L), Medium-Low (ML), Medium (M), Medium-High (MH) and High (H). Each master-copy pair was associated with the level of color similarity expressed by the highest number of observers. This level of color similarity represents the desired output of the network for a considered master-copy pair.

The network has 5 neurons in the output layer, each associated with one level of color similarity among those considered. The output of the network is:

- 0,0,0,0,1 $\rightarrow$ class L;
- 0,0,0,1,0 $\rightarrow$ class ML;
- 0,0,1,0,0 $\rightarrow$ class M;
- 0,1,0,0,0 $\rightarrow$ class MH;
- 1,0,0,0,0 $\rightarrow$ class H.

Fig. 6 shows some examples of assessment by the observers.

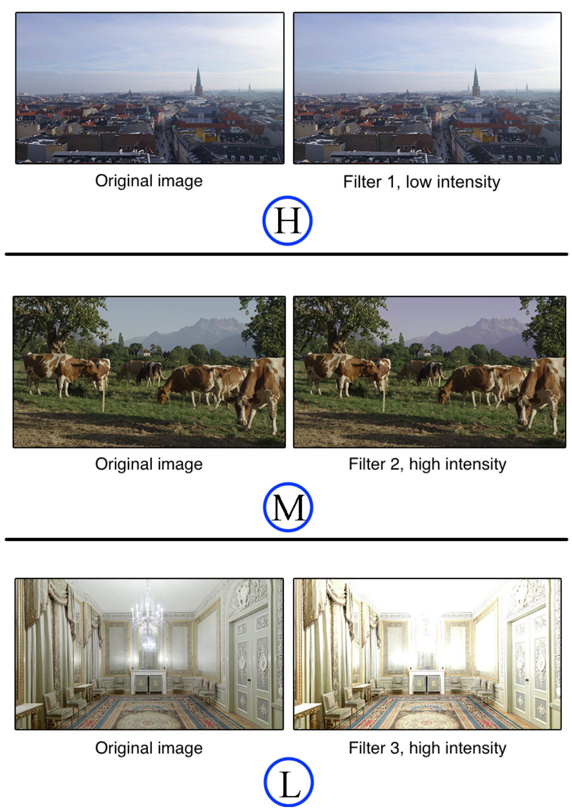

Fig. 6. Levels of color difference expressed by an observer when comparing three master-copy pairs. The level of difference of each pair is inside the blue circle underneath the pair.

V. Results and Discussion

This section first explains how the classifier was designed and trained. The section continues presenting an improved version of the classifier, and discusses the experiments made to assess the role of image resolution in the perception of colors.

A. First version of the classifier

The initial dataset was divided into three subsets: training, validation and test sets, respectively used to train, validate and test the network.

Various networks were trained to find the best architecture. Based on heuristic considerations on the problem, a single hidden layer was used, whose number of neurons was then found by performing a series of experiments where the number of neurons was varied from 5 to 20, with step one. For each number of neurons, the network was trained ten times and the mean cross entropy (MCE) on the test set was calculated. The number of hidden neurons that achieved the lowest MCE was finally chosen. The network was trained by using the Levenberg-Marquardt backpropagation as it showed the best performance.

The confusion matrix of the best classifier obtained is shown in Fig. 7. The network has quite a low precision as it achieves

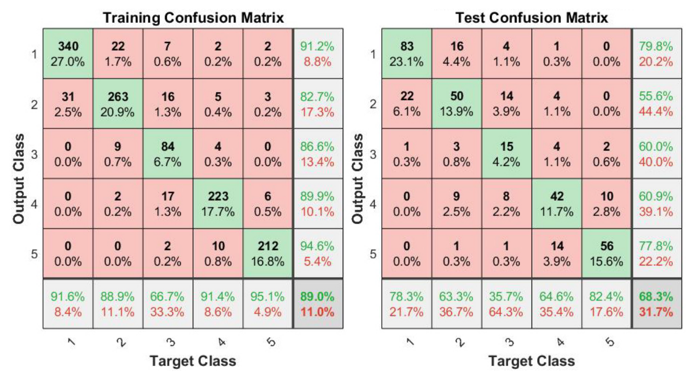

Fig. 7. Confusion matrices of the classifier.

an accuracy just under 70%, precisely 68.3% on the test set. The accuracy can be improved. However, with the exception of one master-copy pair where the target class 5 (high similarity) is confused with class 2 (medium-low similarity), the other classes are confused with levels of color similarity that are mostly one-level far from the level that the network should return.

B. Improving the classification accuracy

1) Feature selection: In this step, the forward sequential feature selection (SFS) algorithm was applied in order to identify the features with the highest discriminating power among those previously extracted.

After 100 iterations, the SFS algorithm selected several features multiple times. The cumulative histogram is shown in Fig. 8. Each column of the histogram is associated with

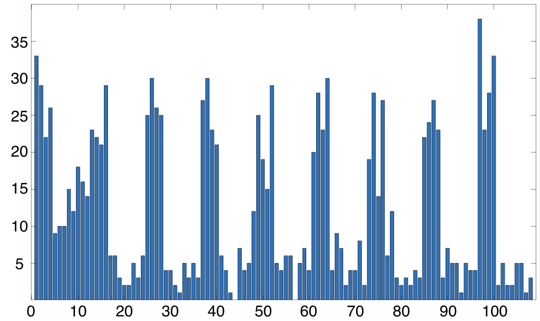

Fig. 8. Cumulative histogram that shows how many times (y-axis) each feature (x-axis) was selected throughout the iterations of the SFS algorithm.

one feature and shows the number of times that the feature—among those in the initial set of 108 features—was chosen throughout the various iterations of the SFS algorithm.

After analyzing the outcome shown in Fig. 8, the most selected features were considered, paying attention to selecting at least one feature for each square of both the original image (X) and the copy (Y). The number of features was thus reduced from 108 to 18.

2) Balancing the dataset: Another cause of non-excellent accuracy of the classifier described in Section V-A might also have been the unbalancing of the dataset. The dataset contains the following number of samples per class:

- class H: 528 samples;
- class MH: 416 samples;
- class M: 183 samples;
- class ML: 349 samples;
- class L: 324 samples.

There is actually an imbalance caused by an excessive number of samples in class H, and a low number of samples in class M. In order to rebalance the dataset, an undersampling was performed to reduce the number of samples of class H. Also, an oversampling was performed in class M (duplication of the samples). The balanced dataset contains a total of 1883 samples.

A new network was then trained, with a more balanced number of samples per class.

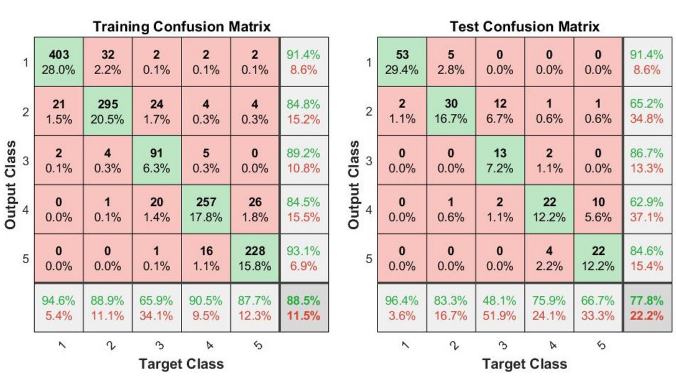

Fig. 9. Confusion matrices of the improved classifier.

As can be seen from the confusion matrices in Fig. 9, the accuracy of the classifier is increased by more than 10%, as it is close to 80%, on the test set. The classes that are predicted with greater precision are those associated with very high and very low similarity. In the intermediate classes, the classifier is less precise. However, the target class is confused with classes that are close to it, thus making low the magnitude of the error.

3) Assessing the role of resolution: The neural network developed was used to understand how the perception of the color difference changes as the resolution decreases. This helps find the minimum resolution at which the images can be acquired in order for the difference in color to remain visible.

Fig. 10 summarizes the analysis that was carried out. In particular, the figure considers 8 master images of the initial dataset and shows how the perceived level of color similarity varies from 1 (i.e., L) to 5 (i.e., H) when comparing a given master image to its copies, at different resolutions.

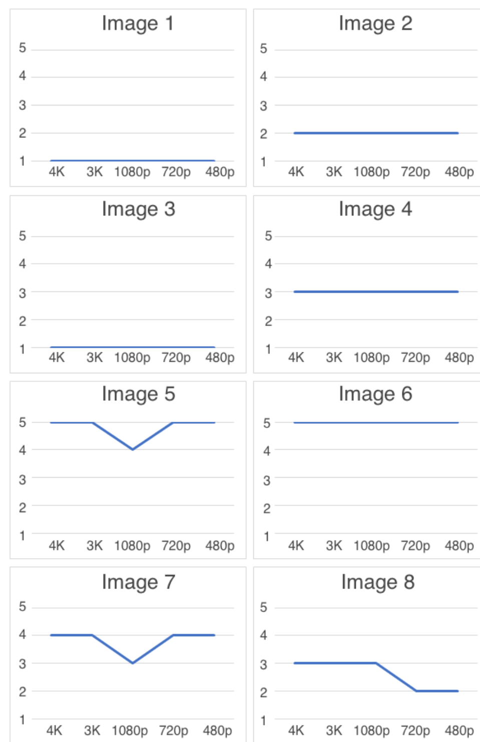

Fig. 10. Level of color similarity from 1 (Low) to 5 (High) perceived by the neural network as the image resolution varies.

Analyzing the results with the help of Fig. 10, it is clear that the color differences between two images are still visible as the resolution decreases, and in some cases they are slightly more evident.

As a consequence, color inspections could be performed by the neural network at 480p, instead of at 4K. If the process were carried out at 4K, a total of 8,294,400 pixels should be processed. On the other hand, at 480p the number of pixels is dramatically reduced to 261,120, with a reduction in number of pixels of $\sim$97%. This result shows that the system could lead to impressive savings of time and money, when dealing with images like those in the dataset. Also, the acquisition of the images could be performed by using less expensive devices that work at lower resolutions. Time savings would also allow to execute a higher number of inspections on many more items, thus increasing the chances of quickly identifying and fixing possible flaws in the production processes.

VI. Conclusions

This paper has presented a neural network to measure the level of color similarity of two images with an accuracy close to 80%. The network was then used to investigate how the perception of the color difference changes as the resolution decreases.

The results highlighted that the color differences continue to be visible, with unchanged level, as the resolution decreases. In some cases, the level of difference is more evident at lower resolutions. The system could thus help improve the efficiency of color inspections which could be performed on more items, at lower resolutions, without human intervention. This could sensibly reduce the costs, while increasing speed and objectivity.

Future works include a system that increases the accuracy of the classifier with the use of AI techniques based on hyperspectral imaging.

Acknowledgment

This work was partially supported by the Italian Ministry of Education and Research (MIUR) in the framework of the CrossLab project (Departments of Excellence), and by the Tuscany Region and Macs Tech S.r.l., under the POR FSE 2014–2020 programme.

References

[1] *https://www.smitherspira.com*. Worldwide authority on the packaging, paper and print industry supply chains.

[2] H. Foley and M. Matlin, *Sensation and Perception*. Psychology Press, 2015.

[3] A. Gilchrist and J. Nobbs, "Colorimetry, theory," in *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)*, third edition ed., J. C. Lindon, G. E. Tranter, and D. W. Koppenaal, Eds. Oxford: Academic Press, 2017, pp. 328 – 333.

[4] A. D. Broadbent, "Colorimetry, methods," in *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)*, third edition ed., J. C. Lindon, G. E. Tranter, and D. W. Koppenaal, Eds. Oxford: Academic Press, 2017, pp. 321 – 327.

[5] S. Billmeyer, *Principles of Color Technology*. New York: Wiley & Sons, 2000.

[6] B. Lazzerini and F. Pistolesi, "A neural system for faithful color reproduction in industrial printing processes," in *2018 International Joint Conference on Neural Networks (IJCNN)*, July 2018, pp. 1–8.

[7] H. Yu, T. Cao, B. Li, R. Dong, and H. Zhou, "A method for color calibration based on simulated annealing optimization," in *2016 3rd International Conference on Information Science and Control Engineering (ICISCE)*, July 2016, pp. 54–58.

[8] Y. Wang, X. Wang, and J. Li, "Color appearance phenomena under high ambient illumination," *Optik - International Journal for Light and Electron Optics*, vol. 145, pp. 22 – 29, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0030402617307878

[9] M. Ataeefard, "Investigating the effect of paper properties on color reproduction of digital printing," *Progress in Organic Coatings*, vol. 77, no. 9, pp. 1376 – 1381, 2014.

[10] H.-J. Kwon, S.-H. Lee, G.-Y. Lee, and K.-I. Sohng, "Luminance adaptation transform based on brightness functions for ldr image reproduction," *Digital Signal Processing*, vol. 30, pp. 74 – 85, 2014.

[11] A. P. Perroni, C. D. Bergoli, M. B. F. dos Santos, R. R. Moraes, and N. Boscato, "Spectrophotometric analysis of clinical factors related to the color of ceramic restorations: A pilot study," *The Journal of Prosthetic Dentistry*, vol. 118, no. 5, pp. 611 – 616, 2017.

[12] J. M. Medina and J. A. Daz, "Classification of batch processes in automotive metallic coatings using principal component analysis similarity factors from reflectance spectra," *Progress in Organic Coatings*, vol. 88, pp. 75 – 83, 2015.

[13] T. Chandrakanth and B. Sandhya, "Quality assessment of images using ssim metric and ciede2000 distance methods in lab color space," in *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*, S. C. Satapathy, B. N. Biswal, S. K. Udgata, and J. K. Mandal, Eds. Cham: Springer International Publishing, 2015, pp. 457–464.