



Article

Comparative Evaluation of Hand-Crafted Image Descriptors vs. Off-the-Shelf CNN-Based Features for Colour Texture Classification under Ideal and Realistic Conditions

Raquel Bello-Cerezo ¹, Francesco Bianconi ^{1,*}, Francesco Di Maria ¹ , Paolo Napoletano ² 
and Fabrizio Smeraldi ³

¹ Department of Engineering, Università degli Studi di Perugia, Via Goffredo Duranti 93–06125 Perugia (PG), Italy; bellocerezo@gmail.com (R.B.-C.); francesco.dimaria@unipg.it (F.D.M.)

² Department of Informatics, Systems and Communication, Università degli studi di Milano-Bicocca, Viale Sarca 336–20125 Milano (MI), Italy; paolo.napoletano@disco.unimib.it

³ School of Electronic Engineering and Computer Science, Queen Mary University of London, 10 Godward Square, Mile End Road, London E1 4FZ, UK; f.smeraldi@qmul.ac.uk

* Correspondence: bianco@ieee.org; Tel.: +39-075-5853706

Received: 16 January 2019; Accepted: 15 February 2019; Published: 20 February 2019



Abstract: Convolutional Neural Networks (CNN) have brought spectacular improvements in several fields of machine vision including object, scene and face recognition. Nonetheless, the impact of this new paradigm on the classification of fine-grained images—such as colour textures—is still controversial. In this work, we evaluate the effectiveness of traditional, hand-crafted descriptors against off-the-shelf CNN-based features for the classification of different types of colour textures under a range of imaging conditions. The study covers 68 image descriptors (35 hand-crafted and 33 CNN-based) and 46 compilations of 23 colour texture datasets divided into 10 experimental conditions. On average, the results indicate a marked superiority of deep networks, particularly with non-stationary textures and in the presence of multiple changes in the acquisition conditions. By contrast, hand-crafted descriptors were better at discriminating stationary textures under steady imaging conditions and proved more robust than CNN-based features to image rotation.

Keywords: colour texture; feature extraction; image classification; convolutional neural networks; hand-crafted image descriptors

1. Introduction

Colour texture analysis and classification play a pivotal role in many computer-vision applications such as surface inspection, remote sensing, medical image analysis, object recognition, content-based image retrieval and many others. It is no surprise, then, that texture has been an area of intense research activity for at least forty years and that a huge variety of descriptors has been proposed in the literature (see [1–3] for comprehensive and up-to-date reviews).

In recent years, the advent of convolutional neural networks has dramatically changed the outlook in many areas of computer vision and has led to astonishing improvements in tasks like object, scene and face recognition [4–8]. The structure of a CNN differs from that of traditional, hand-designed descriptors in that the former contains a large number of parameters, which are to be determined through suitable training procedures. The training process is usually carried out on huge datasets (containing millions of images), which enables the nets to “learn” very complex image-to-feature and/or image-to-class mappings. More importantly, there is evidence that such

mappings are amenable to being transferred from one domain to another, making networks trained on certain classes of images usable in completely different contexts [5,9]. The consequences of this are far-reaching: although datasets large enough to train a CNN entirely from scratch are rarely available in practical tasks, pre-trained networks can in principle be used as off-the-shelf feature extractors in a wide range of applications.

Inevitably, the CNN paradigm is changing the approach to texture analysis as well. However, though there is general consensus that convolutional networks are superior to hand-designed descriptors in tasks such as object and scene recognition [4–6], this superiority is still not quite clear when it comes to dealing with *fine-grained* images, i.e., textures. As we discuss in Section 2, some recent results seem to point in that direction, but it is precisely the aim of this work to investigate this matter further. To this end, we comparatively evaluated the performance of a large number of classic and more recent hand-designed, local image descriptors against a selection of off-the-shelf features from last-generation CNN. We assessed the performance under both ideal and realistic conditions, with special regard to different degrees of intra-class variability. As we detail in Section 2, intra-class variability can be the consequence of the intrinsic structure of the texture—which can be more or less stationary—and/or of variations in the imaging conditions (e.g., changes in illumination, rotation, scale and/or viewpoint).

On the whole, image features from pre-trained networks outperformed hand-crafted descriptors, but with some interesting exceptions. In particular, hand-crafted methods were still better than CNN-based features at discriminating between very similar colour textures under invariable imaging conditions and proved slightly more robust to rotation; whereas the results were split in the presence of variations of scale.

Networks were markedly superior in all the other cases—particularly with non-stationary colour textures—and also emerged as more robust to multiple and uncontrolled changes in the imaging conditions, which are harder to model and compensate for in a priori feature design.

In the remainder of the paper, we first put the work in the context of the recent literature (Section 2), then describe the datasets and image descriptors used in our study (Sections 3–4). We detail the experimental set-up in Section 5, discuss the results in Section 6 and conclude the paper with some final considerations in Section 7.

2. Related Research

A number of papers have addressed the problem of comparatively evaluating image descriptors for colour texture analysis (e.g., [10–14]), though none of these considers CNN-based features. Yet, results from recent studies seem to suggest that CNN-based descriptors can be effective for texture classification as well, in most cases outperforming hand-designed descriptors.

Cimpoi et al. [15,16] compared a number of hand-designed image descriptors including LMfilters, MR8filters, LBP and SIFT against a set of CNN-based features in texture and material recognition and concluded that in most cases, the latter group outperformed the former. Notably, their findings are mainly based on the results obtained on the Describable Texture Dataset (DTD; more on this in Section 3), which to a great extent, is composed of very irregular and non-stationary textures acquired “in the wild”; by contrast, their results look rather saturated and levelled in other datasets (i.e., Columbia-Utrecht Reflectance and Texture Database (CUReT), UMDand UIUC). Cusano et al. [17] investigated colour texture classification under variable illumination conditions. Their study included a large number of hand-designed texture descriptors (e.g., Gabor filters, wavelet and LBP), descriptors for object recognition (dense SIFT) and CNN-based features. Experimenting on a new dataset of colour texture images (RawFooT, also included in the present study), they concluded that features based on CNN gave significantly better results than the other methods. Liu et al. [1] evaluated a large selection of LBP variants and CNN-based features for texture classification tasks. In their experiments, CNN-based features outperformed LBP variants in six datasets out of eleven, but in this case, all the LBP variants considered were grey-scale descriptors, whereas CNN by default operates on colour

images. The presence/absence of colour information may account—at least in part—for the different performance. Recently, Napoletano [18] comparatively evaluated a number of hand-crafted descriptors and CNN-based features over five datasets of colour images and found that, on average, the latter group outperformed the former.

In summary, there is mounting evidence that off-the-shelf CNN-based features can be suitable for texture classification tasks and may in certain cases outperform (therefore, potentially replace) traditional, hand-designed descriptors. Interestingly, both [1,17] seem to suggest that CNN-based methods tend to perform better than hand-designed descriptors in the presence of complex textures and intra-class variations, though none of the references investigated this point further. There remains a need to clarify under what circumstances CNN-based features can replace traditional, hand-crafted descriptors and what are the pros and cons of the two strategies.

3. Materials

We based our experiments on 23 datasets of colour texture images (Section 3.1) arranged into 46 different experimental conditions (Section 3.2). We subdivided the datasets into ten different groups (Sections 3.2.1–3.2.10) based on the following two properties of the images contained therein (see also Table 1 and Figures 1–10):

- (a) The stationariness of the textures;
- (b) The presence/absence and/or the type of variation in the imaging conditions.

Table 1. Round-up table of the image datasets used in the experiments. ALOT, Amsterdam Library of Textures; CURET, Columbia-Utrecht Reflectance and Texture Database; CBT, Coloured Brodatz Textures; MBT, Multiband Texture Database; RawFooT, Raw Food Texture Database; VisTex, Vision Texture Database; RDAD, Robotics Domain Attribute Database; STex, Salzburg Texture Image Database; DTD, Describable Texture Dataset; S, Stationary; NS, Non-Stationary; N, No variations; I, variations in Illumination; R, variations in Rotation; S, variations in Scale, M, Multiple variations.

		VARIATIONS IN THE IMAGING CONDITIONS				
		NONE Group 1	ILLUMINATION Group 3	ROTATION Group 5	SCALE Group 7	MULTIPLE VARIATIONS Group 9
TYPE OF TEXTURE	STATIONARY	ALOT-95-S-N	ALOT-95-S-I	ALOT-95-S-R	KTH-TIPS-10-S-S	CURET-61-S-M
	CBT-99-S-N	Outex-192-S-I	KylbergSintorn-25-S-R	KTH-TIPS2b-11-S-S	Fabrics-1968-S-M	
	Drexel-18-S-N	RawFooT-68-S-I1	MondialMarmi-25-S-R	Outex-192-S-S	KTH-TIPS-10-S-M	
	KylbergSintorn-25-S-N	RawFooT-68-S-I2	Outex-192-S-R		KTH-TIPS2b-11-S-M	
	MBT-120-S-N	RawFooT-68-S-I3			LMT-94-S-M	
	MondialMarmi-25-S-N				RDAD-27-S-M	
	Outex-192-S-N					
	Parquet-38-S-N					
	PlantLeaves-20-S-N					
	RawFooT-68-S-N					
	STex-202-S-N					
	USPTex-137-S-N					
	VisTex-89-S-N					
	VxC_TSG-42-S-N					
		Group 2	Group 4	Group 6	Group 8	Group 10
NON-STATIONARY	ALOT-40-NS-N	ALOT-40-NS-I	ALOT-40-NS-R	Outex-59-NS-S	DTD-47-NS-M	
	ForestSpecies-112-NS-N	Outex-59-NS-I	Outex-59-NS-R			
	MBT-34-NS-N					
	NewBarkTex-6-NS-N					
	Outex-59-NS-N					
	STex-138-NS-N					
	USPTex-33-NS-N					
	VisTex-78-NS-N					

Property (a) refers to the concept of *stationariness* as defined in [19], i.e., a texture that “fills up the whole image and its local statistical properties are the same everywhere in it”. In this case, the subdivision is binary, which means we can have either *stationary* or *non-stationary* textures. Property

(b) signals whether the samples of a given class have been acquired under steady or variable imaging conditions in terms of illumination, rotation, scale and/or viewpoint. In the remainder, we use the following naming convention to indicate the dataset used:

`<source>-<no.-of-classes>-<prop-a>-<prop-b>`

where:

- `<source>` indicates the name of the source dataset the images come from (e.g., Amsterdam Library of Textures (ALOT), KTH-TIPS, etc. as detailed in Section 3.1);
- `<no.-of-classes>` the number of the colour texture classes in the dataset;
- `<prop-a>` the stationariness of the textures, which can be either S or NS respectively indicating *Stationary* and *Non-stationary* textures;
- `<prop-b>` the presence/absence and/or the type of intra-class variation in the imaging conditions. This can be either N, I, R, S or M, respectively indicating No variations (steady imaging conditions), variations in *Illumination*, variations in *Rotation*, variations in *Scale* and *Multiple variations* (i.e., combined changes in illumination, scale, rotation and/or viewpoint).

3.1. Source Datasets

3.1.1. Amsterdam Library of Textures

This is a collection of stationary and non-stationary colour textures representing 250 classes of heterogeneous materials including *chip*, *fabric*, *pebble*, *plastics*, *seeds* and *vegetables* [20,21]. Each class was acquired under 100 different conditions obtained by varying the viewing direction, the illumination direction and the rotation angle. The dataset comes in full, half or quarter resolution (respectively 1536 px × 1024 px, 768 px × 512 px and 384 px × 256 px): we chose the first for our experiments.

3.1.2. Coloured Brodatz Textures

Coloured Brodatz Textures (CBT) is an artificially-colourised version of Brodatz's album [22,23]. There are 112 classes with one image sample for each class. The dimension of the images is 640 px × 640 px, which we subdivided into four non-overlapping sub-images of dimensions 320 px × 320 px.

3.1.3. Columbia-Utrecht Reflectance and Texture Database

The Columbia-Utrecht Reflectance and Texture Database (CURET) contains sixty-one classes representing different types of materials such as *aluminium foil*, *artificial grass*, *brick*, *cork*, *cotton*, *leather*, *quarry tile*, *paper*, *sandpaper*, *styrofoam* and *velvet* [24,25]. In the original version, there are 205 image samples for each class corresponding to different combinations of viewing and illumination directions. Some of these images, however, cannot be used because they contain only a small portion of texture, while the rest is background. The version used here is a reduced one [26] maintained by the Visual Geometry Group at the University of Oxford, United Kingdom. In this case, there are 92 images per class corresponding to those imaging conditions that ensure a sufficiently large texture portion to be visible across all materials. The dimension of each image sample is 200 px × 200 px.

3.1.4. Drexel Texture Database

This consists of stationary colour textures representing 20 different materials such as *bark*, *carpet*, *cloth*, *knit*, *sandpaper*, *sole*, *sponge* and *toast* [27,28]. The dataset features 1560 images per class, which are the result of combining 30 viewing conditions (generated by varying the object-camera distance and the angle between the camera axis and the imaged surface) and 52 illumination directions. The images have a dimension of 128 px × 128 px.

3.1.5. Describable Textures Dataset

DTD is comprised of highly non-stationary and irregular textures acquired under uncontrolled imaging conditions (or, as the authors say, “in the wild” [29,30]). The images are grouped into 47 classes representing attributes related to human perception such as *banded, blotchy, cracked, crystalline, dotted, meshed* and so forth. There are 120 samples per class, and the dimension of the images varies between 300 px × 300 px and 640 px × 640 px.

3.1.6. Fabrics Dataset

This is comprised of 1968 samples of garments and fabrics [31,32]. Herein, we considered each sample as a class on its own, though the samples are also grouped by material (e.g., *wool, cotton, polyester*, etc.) and garment type (e.g., *pants, shirt, skirt* and the like). The images were acquired in the field (i.e., at garment shops) using a portable photometric device and have a dimension of 400 px × 400 px. Each sample was acquired under four different illumination conditions; the samples also have uncontrolled in-plane rotation.

3.1.7. Forest Species Database

The Forest Species Database (ForestSpecies) is comprised of 2240 images representing samples from 112 hardwood and softwood species [33,34]. The images were acquired through a light microscope with 100× zoom and have a dimension of 1024 px × 768 px.

3.1.8. KTH-TIPS

This is comprised of ten types of materials such as *aluminium foil, bread, cotton* and *sponge* [35,36]. Each material sample was acquired under nine different scales, three rotation angles and three illumination directions, giving 81 images for each class. The dimension of the images is 200 px × 200 px.

3.1.9. KTH-TIPS2b

KTH-TIPS2bis an extension to KTH-TIPS, which adds one more class, three more samples per class and one additional illumination condition [36,37]. As a result, there are 432 image samples per class instead of 81, whereas the image dimension is the same as in KTH-TIPS.

3.1.10. Kylberg–Sintorn Rotation Dataset (*KylbergSintorn*)

The Kylberg–Sintorn Rotation Dataset (*KylbergSintorn*) is comprised of twenty-five colour texture classes representing common materials such as *sugar, knitwear, rice, tiles* and *wool* [38,39]. There is one sample for each class, which was acquired at nine in-plane rotation angles, i.e., 0°, 40°, 80°, 120°, 160°, 200°, 240°, 280° and 320°. The dimension of the images is 5184 px × 3456 px.

3.1.11. LMT Haptic Texture Database

LMTis comprised of one hundred eight colour texture classes belonging to the following nine super-classes: *blank glossy surfaces, fibres, foams, foils and papers, meshes, rubbers, stones, textiles and fabrics* and *wood* [40,41]. The images were acquired using a common smartphone, and each material sample was captured under 40 different illumination and viewing conditions. The dimension of the images is 320 px × 480 px.

3.1.12. MondialMarmi

This is comprised of twenty-five classes of commercial polished stones [42,43]. Four samples per class (corresponding to as many tiles) were acquired under steady and controlled illumination conditions and at 10 in-plane rotation angles, i.e., 0°, 10°, 20°, 30°, 40°, 50°, 60°, 70°, 80° and 90°. The dimension of the images is 1500 px × 1500 px.

3.1.13. Multiband Texture Database

The Multiband Texture Database (MBT) is comprised of one hundred fifty-four colour texture images obtained by taking three grey-scale textures [44,45] taken from the Normalized Brodatz Texture database [46] and dealing with one of each of the R, G and B channels. There is one sample for each class, and the image size is 640 px × 640 px.

3.1.14. New BarkTex

BarkTex is a collage of different types of tree bark [47,48] derived from the BarkTex database [49]. This dataset includes six classes with 68 samples per class. The dimension of the images is 64 px × 64 px.

3.1.15. Outex Texture Database

Outex is a well-known collection of texture images from 319 diverse materials such as *canvas*, *cardboard*, *granite*, *leather*, *seeds*, *pasta* and *wood* [50,51]. There are 162 images for each class resulting from acquiring the corresponding material under six different scales (100 dpi, 120 dpi, 300 dpi, 360 dpi, 500 dpi and 600 dpi), nine in-plane rotation angles (0°, 5°, 10°, 15°, 30°, 45°, 60°, 75° and 90°) and three illuminants (“inca”, “TL84” and “horizon”). All the images have a dimension of 746 px × 538 px.

3.1.16. Parquet

This is comprised of fourteen commercial varieties of finished wood for flooring and cladding [52,53]. Each variety has also a number of grades ranging from 2–4, which we considered as independent classes, yielding a total of 38 classes. The number of samples per class varies from 6–8 and the dimension of the images from 1200 px–1600 px in width and from 500 px–1300 px in height, as a consequence of the different sizes of the specimens.

3.1.17. Plant Leaves Database

This is comprised of twenty classes of plant leaves from as many plant species [54,55]. Three images of dimensions 128 px × 128 px were acquired from the regions of minimum texture variance within each leaf, making a total of 20 × 20 × 3 = 1200 images. The acquisition was carried out using a planar scanner at a spatial resolution of 1200 dpi.

3.1.18. Robotics Domain Attribute Database

Robotics Domain Attribute Database (RDAD) is comprised of fifty-seven classes of objects and materials such as *asphalt*, *chocolate*, *coconut*, *flakes*, *pavingstone*, *rice* and *styrofoam* [56]. The dataset includes a variable number of image samples per class (from 20–48), all captured “in the wild”. The dimension of each image is 2592 px × 1944 px.

3.1.19. Raw Food Texture Database

The Raw Food Texture Database (RawFooT) is comprised of sixty-eight classes of raw food such as *chickpeas*, *green peas*, *oat*, *chilly pepper*, *kiwi*, *mango*, *salmon* and *sugar* [57,58]. Each image was taken under 46 different illumination conditions obtained by varying the type, the direction and the intensity of the illuminant; other imaging conditions such as scale, rotation and viewpoint remained invariable. The images have a dimension of 800 px × 800 px.

3.1.20. Salzburg Texture Image Database

The Salzburg Texture Image Database (STex) is comprised of four hundred seventy-six colour texture images acquired “in the wild” around the city of Salzburg, Austria [59]. They mainly represent objects and materials like *bark*, *floor*, *leather*, *marble*, *stones*, *walls* and *wood*. The dataset comes in two different resolutions—i.e., 1024 px × 1024 px and 512 px × 512 px—of which the second was the

one used in our experiments. We further subdivided the original images into 16 non-overlapping sub-images of dimensions $128 \text{ px} \times 128 \text{ px}$.

3.1.21. USPTex

USPTex [60,61] is very similar to STex (Section 3.1.20) as for the content, structure and imaging conditions. In this case, there are 191 classes representing materials, objects and scenes such as *food, foliage, gravel, tiles* and *vegetation*. There are 12 samples per class and the image dimensions $128 \text{ px} \times 128 \text{ px}$.

3.1.22. VisTex Reference Textures

The VisTex reference textures are part of the Vision Texture Database [62]. They represent 167 classes, which are further subdivided into 19 groups, e.g., *bark, buildings, food, leaves, terrain* and *wood*. For each class, there is one image sample of dimensions $512 \text{ px} \times 512 \text{ px}$, which we subdivided into four non-overlapping samples of $256 \text{ px} \times 256 \text{ px}$.

3.1.23. VxC TSG Database

VxCTSG is comprised of fourteen commercial classes of ceramic tiles with three grades per class [63]. We considered each grade as a class on its own, which gives 42 classes in total. The images were acquired in a laboratory under controlled and invariable conditions. The number of samples per class varies from 14–30, but in our experiments, we only retained 12 samples per class. Since the original images are rectangular, we cropped them to a square shape, retaining the central part. The resulting images have a dimension ranging between $500 \text{ px} \times 500 \text{ px}$ and $950 \text{ px} \times 950 \text{ px}$.

3.2. Datasets Used in the Experiments

From the source datasets (Section 3.1), we derived the datasets used in the experiments. The classification into stationary or non-stationary textures was performed manually by two of the authors (R.B.-C, >2 years experience in texture analysis, and F.B., >10 years experience in texture analysis). Those images on which no consensus was reached were discarded.

3.2.1. Group #1: Stationary Textures Acquired under Steady Imaging Conditions

- ALOT-95-S-N: Ninety-five stationary textures from the ALOT dataset (Section 3.1.1) with images taken from the “c1I3” group. Six samples per class were obtained by subdividing the original images into non-overlapping sub-images of dimensions $256 \text{ px} \times 256 \text{ px}$.
- CBT-99-S-N: Ninety-nine stationary textures from the CBT dataset (Section 3.1.2).
- Drexel-18-S-N: Eighteen stationary textures from the Drexel dataset (Section 3.1.4) with images taken from the “D1_IN00_OUT00” group. Four samples per class were obtained by subdividing the original images into non-overlapping sub-images of dimensions $64 \text{ px} \times 64 \text{ px}$.
- KylbergSintorn-25-S-N: All 25 colour textures in the KylbergSintorn dataset (Section 3.1.10) with images taken from the 0° group. Each image was subdivided into 24 non-overlapping images of dimensions $864 \text{ px} \times 864 \text{ px}$.
- MBT-120-S-N: One hundred and twenty stationary colour textures from MBT (Section 3.1.13). Each image was subdivided into four non-overlapping sub-images of dimensions $320 \text{ px} \times 320 \text{ px}$, giving four samples per class.
- MondialMarmi-25-S-N: All 25 colour textures of the MondialMarmi dataset (Section 3.1.12) with images taken from the 0° group. Each image was subdivided into four non-overlapping sub-images of dimensions $750 \text{ px} \times 750 \text{ px}$, giving 16 samples per class.
- Outex-192-S-N: One hundred ninety-two stationary colour textures from the Outex dataset (Section 3.1.15) with images acquired under the following conditions: scale = 100 dpi, rotation 0° and illuminant = “inca”. We cropped the central part of each image and subdivided it into 20 non-overlapping sub-images of dimensions $128 \text{ px} \times 128 \text{ px}$.

- Parquet-37-S-N: All 38 classes of the Parquet dataset (Section 3.1.16). We retained six samples per class and centre-cropped the images. The final dimension ranges from 480 px × 480 px–1300 px × 1300 px.
- PlantLeaves-20-S-N: The entirety of the PlantLeaves dataset (Section 3.1.17).
- RawFoot-68-S-N: All 68 classes of the RawFoot dataset (Section 3.1.19) with the following imaging conditions: illuminant = “D65” and illumination intensity = 100%. We obtained four samples per class by subdividing the original images into four non-overlapping tiles of dimensions 400 px × 400 px.
- STex-202-S-N: Two hundred and two stationary colour textures from STex (Section 3.1.20).
- USPTex-137-S-N: One hundred thirty-seven stationary colour textures from USPTex (Section 3.1.21).
- VisTex-89-S-N: Eighty-nine stationary colour textures from VisTex (Section 3.1.22). Each image was subdivided into four non-overlapping sub-images of dimensions 256 px × 256 px.
- VxC_TSG-42-S-N: All 42 classes of the VxC_TSG dataset (Section 3.1.23).

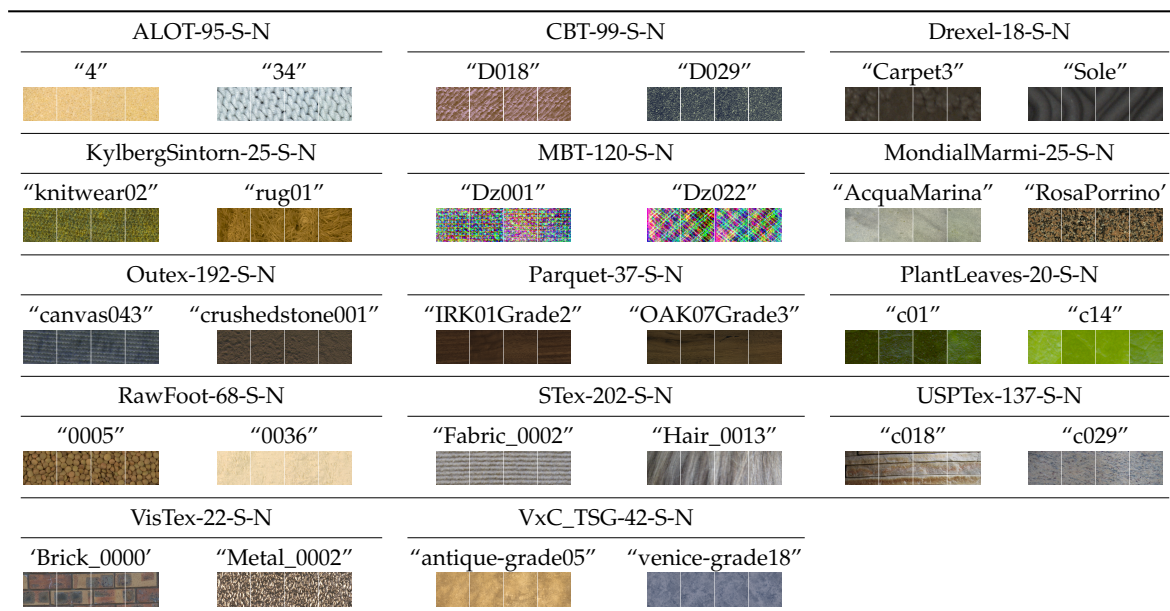


Figure 1. Group #1: Stationary textures acquired under steady imaging conditions (Section 3.2.1).

3.2.2. Group #2: Non-Stationary Textures Acquired under Steady Imaging Conditions

- ALOT-40-NS-N: Forty non-stationary colour textures from the ALOT dataset (Section 3.1.1). The other settings were the same as in ALOT-95-S-N.
- ForestSpecies-112-NS-N: The entirety of the ForestSpecies dataset (Section 3.1.7).
- MBT-34-NS-N: Thirty-four non-stationary colour textures from the MBT dataset (Section 3.1.7). The other settings were the same as in the MBT-120-S-N dataset.
- NewBarkTex-6-NS-N: The whole New BarkTex dataset (Section 3.1.14).
- Outex-59-NS-N: Fifty-nine non-stationary colour textures from the Outex dataset (Section 3.1.15). The other settings were the same as in Outex-192-S-N.
- STex-138-NS-N: One hundred thirty-eight non-stationary colour textures from the STex dataset (Section 3.1.20). The other settings were the same as in STex-202-S-N.
- USPTex-33-NS-N: Thirty-three non-stationary colour textures from the USPTex dataset (Section 3.1.21). The other settings were the same as in USPTex-137-S-N.
- VisTex-78-NS-N: Seventy-eight non-stationary textures from the VisTex dataset (Section 3.1.22). The other settings were the same as in VisTex-89-S-N.

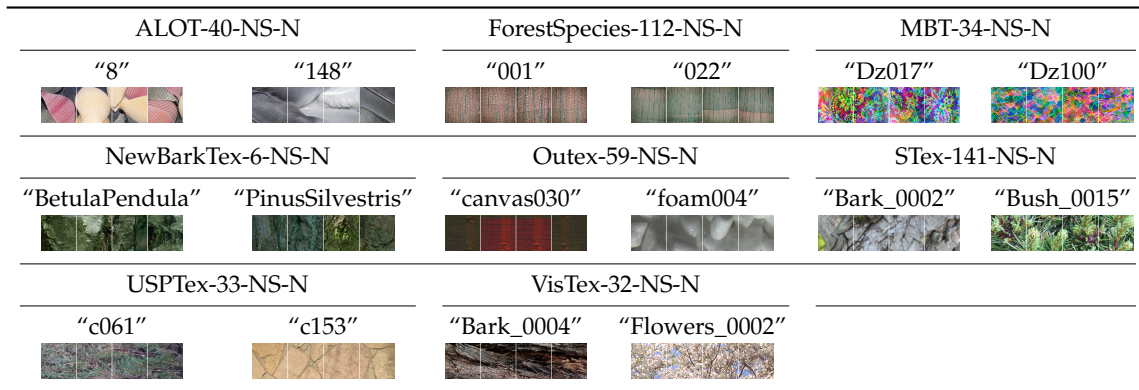


Figure 2. Group #2: Non-stationary textures acquired under steady imaging conditions (Section 3.2.2).

3.2.3. Group #3: Stationary Textures Acquired under Variable Illumination

- ALOT-95-S-I: The same ninety-five textures of the ALOT-95-S-N dataset acquired at 0° rotation, orthogonal shot (camera “c1”) and five different illumination directions (conditions “I1”, “I2”, “I3”, “I4” and “I5”). Each image was subdivided into six sub-images as in ALOT-95-S-N, giving $6 \times 5 = 30$ samples per class.
- Outex-192-S-I: The same one hundred ninety-two textures of the Outex-192-S-N dataset acquired at 0°, 100 dpi scale and three different illuminants (i.e., “inca”, “TL84” and “horizon”). As in Outex-192-S-N, each image was subdivided into 20 non-overlapping images, resulting in a total of $20 \times 3 = 60$ samples per class.
- RawFooT-68-S-I1: The same sixty-eight textures of the RawFooT-68-S-N dataset acquired under an invariable light source (D65), but four different intensities, namely: 100%, 75%, 50% and 25%, which respectively correspond to conditions “01”, “02”, “03” and “04” of the RawFooT database. As in RawFooT-68-S-N, there are four samples for each imaging condition, therefore a total of $4 \times 4 = 16$ samples per class.
- RawFooT-68-S-I2: The same sixty-eight textures of the RawFooT-68-S-N dataset acquired under six different light sources, i.e., D40, D55, D70, D85, L27 and L5, respectively corresponding to conditions “14”, “17”, “20”, “23”, “26” and “29” of the RawFooT database. As in RawFooT-68-S-N, there are four samples for each imaging condition, therefore a total of $6 \times 4 = 24$ samples per class.
- RawFooT-68-S-I3: The same sixty-eight textures of the RawFooT-68-S-N dataset acquired under an invariable light source (D65) coming from different illumination directions: $\theta = 24^\circ$, $\theta = 42^\circ$ and $\theta = 60^\circ$, which respectively correspond to conditions “05”, “08” and “11” of the RawFooT database. In this case, there are $3 \times 4 = 12$ samples per class.

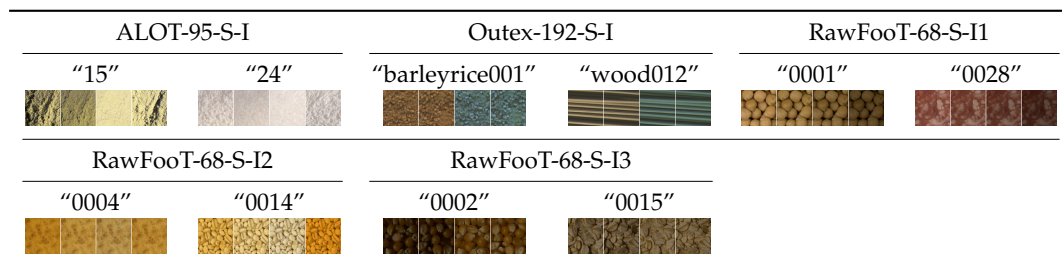


Figure 3. Group #3: Stationary textures acquired under variable illumination (Section 3.2.3).

3.2.4. Group #4: Non-Stationary Textures Acquired under Variable Illumination

- ALOT-40-NS-I: The same forty textures of ALOT-40-S-N and the same acquisition conditions and number of samples per class as in ALOT-95-S-I.

- Outex-59-NS-I: The same fifty-nine textures of Outex-59-S-N and the same acquisition conditions and number of samples per class as in Outex-192-S-I.



Figure 4. Group #4: Non-stationary textures acquired under variable illumination (Section 3.2.4).

3.2.5. Group #5: Stationary Textures with Rotation

- ALOT-95-S-R: The same ninety-five textures of the ALOT-95-S-N dataset acquired with camera “c1”, illumination direction “I3” and four in-plane rotation angles, i.e., 0°, 60°, 120° and 180°. Each image was subdivided into six sub-images as in ALOT-95-S-N, giving $6 \times 4 = 24$ samples per class.
- Outex-192-S-R: The same one hundred ninety-two textures of the Outex-192-S-N dataset acquired with lamp type “inca”, 100 dpi scale and and four in-plane rotation angles, i.e., 0°, 30°, 60° and 90°. Each image was subdivided into 20 non-overlapping images as in Outex-192-S-R, giving a total $20 \times 4 = 80$ samples per class.
- MondialMarmi-25-S-R: The same twenty-five textures of the MondialMarmi-25-S-N dataset acquired under four in-plane rotation angles, i.e., 0°, 30°, 60° and 90°. There are 16 samples per rotation angle as in MondialMarmi-25-S-N, for a total of $16 \times 4 = 64$ samples per class.
- KylbergSintorn-25-S-R: The same twenty-five textures of the KylbergSintorn-25-N-R dataset acquired under four in-plane rotation angles, which in this case were: 0°, 120°, 240° and 320°. The number of samples for each orientation is 24 as in KylbergSintorn-25-N-R, giving a total of $24 \times 4 = 96$ samples per class.

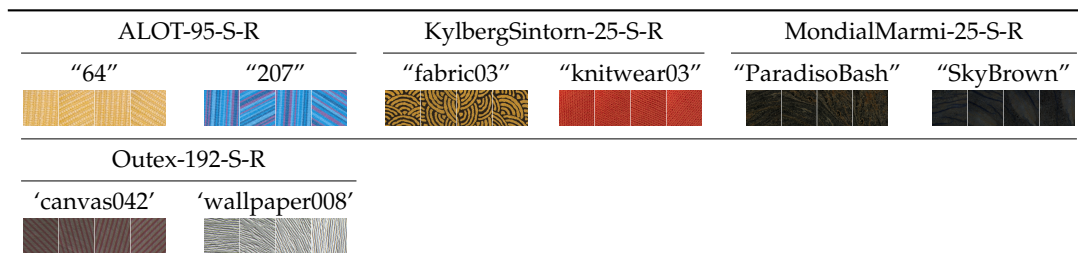


Figure 5. Group #5: Stationary textures with rotation (Section 3.2.5).

3.2.6. Group #6: Non-Stationary Textures with Rotation

- ALOT-40-NS-R: The same forty textures of ALOT-40-NS-N and the same acquisition conditions and number of samples per class as in ALOT-95-S-R.
- Outex-59-NS-R: The same fifty-nine textures as in Outex-59-NS-N the same acquisition conditions and number of samples per class as in Outex-192-S-R.

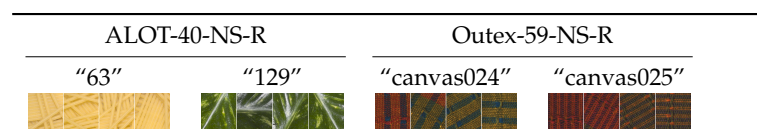


Figure 6. Group #6: Non-stationary textures with rotation (Section 3.2.6).

3.2.7. Group #7: Stationary Textures with Variations in Scale

- KTH-TIPS-10-S-S: All the classes of the KTH-TIPS dataset with nine image samples per class, each sample being taken under fixed pose and illumination (frontal), and nine different relative scales.
- KTH-TIPS2b-11-S-S: Same settings as KTH-TIPS-10-S-S, i.e., fixed pose and illumination (frontal), but variable scale. The number of samples per class is $4 \times 9 = 36$ in this case, for there are four specimens per class.
- Outex-192-S-S: The same one hundred ninety-two textures of Outex-192-S-N taken under fixed illumination (“inca”) and rotation (0°) but variable scale, respectively 100 dpi, 120 dpi, 300 dpi and 600 dpi.

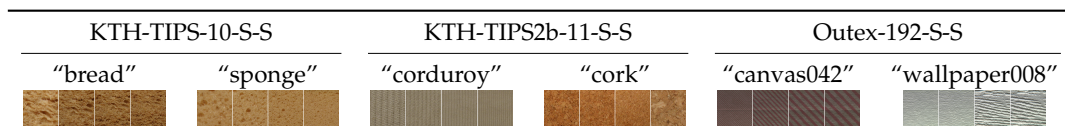


Figure 7. Group #7: Stationary textures with variations in scale (Section 3.2.7).

3.2.8. Group #8: Non-Stationary Textures with Variations in Scale

- Outex-59-S-S: The same fifty-nine textures of Outex-59-S-N taken under the same conditions as in Outex-192-S-S.

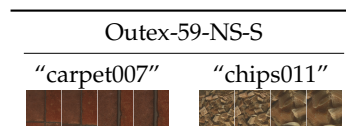


Figure 8. Group #8: Non-stationary textures with variations in scale (Section 3.2.8).

3.2.9. Group #9: Stationary Textures Acquired under Multiple Variations in the Imaging Conditions

- CURET-61-S-M: The complete CURET dataset as described in Section 3.1.3.
- Fabrics-1968-S-M: All 1968 colour textures of the Fabrics dataset (Section 3.1.6).
- KTH-TIPS-10-S-M: The entirety of the KTH-TIPS dataset (Section 3.1.8).
- KTH-TIPS2b-11-S-M: The whole KTH-TIPS2b dataset (Section 3.1.9).
- LMT-94-S-M: Ninety-four stationary textures from the LMT dataset (Section 3.1.10).
- RDAD-27-S-M: A selection of 27 stationary colour textures from the RDAD database (Section 3.1.18). We included the classes representing textures, discarded those representing objects and retained 16 images samples for each class.

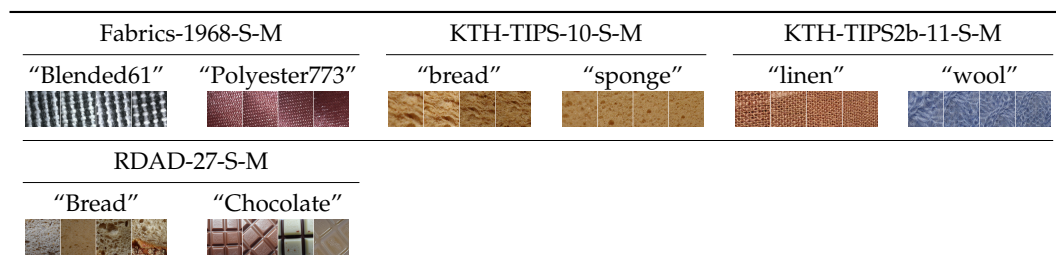


Figure 9. Group #9: Stationary textures acquired under multiple variations in the imaging conditions (Section 3.2.9).

3.2.10. Group #10: Non-Stationary Textures Acquired under Multiple Variations in the Imaging Conditions

- DTD-47-NS-M: The entirety of the DTD dataset (Section 3.1.5).



Figure 10. Group #10: Non-stationary textures acquired under multiple variations in the imaging conditions (Section 3.2.10).

4. Methods

We considered 35 hand-designed and 33 CNN-based descriptors as detailed in Sections 4.1–4.2 (see also Tables 2 and 3 for a round-up). We subdivided the hand-designed methods into three subgroups:

- Purely spectral descriptors (colour descriptors; Section 4.1.1);
- Grey-scale texture descriptors (Section 4.1.2);
- Colour texture descriptors (Section 4.1.3).

Table 2. Summary table of the hand-crafted image descriptors used in the experiments. VLAD, Vectors of Locally-Aggregated Descriptors.

Method	Variant	Abbreviation	No. of Features
<i>Purely spectral descriptors</i>			
Mean of each channel		Mean	3
Mean and std. dev.of each channel		Mean + Std	6
Mean and moms.from 2th to 5th of each ch.		Mean + Moms	15
Quartiles of each channel		Quartiles	9
256-bin Marginal Histogram of each channel		Marginal-Hists-256	768
10-bin joint colour Histogram		Full-Hist-10	1000
<i>Grey-scale texture descriptors</i>			
Completed Local Binary Patterns	Rotation-invariant	CLBP	324
Gradient-based Local Binary Patterns	Rotation-invariant	GLBP	108
Improved Local Binary Patterns	Rotation-invariant	ILBP	213
Local Binary Patterns	Rotation-invariant	LBP	108
Local Ternary Patterns	Rotation-invariant	LTP	216
Texture Spectrum	Rotation-invariant	TS	2502
Grey-level Co-occurrence Matrices		GLCM	60
Grey-level Co-occurrence Matrices		GLCM ^{DFT}	60
Gabor features		Gabor	70
Gabor features	Rotation-invariant	Gabor ^{DFT}	70
Gabor features	Contrast-normalised	Gabor _{cn}	70
Gabor features	Rot.-inv.and ctr.-norm.	Gabor _{cn} ^{DFT}	70
Image Patch-Based Classifier	Joint	IPBC-J	4096
Histograms of Oriented Gradients		HOG	768
Dense SIFT	BoVW aggregation	SIFT-BoVW	4096
Dense SIFT	VLAD aggregation	SIFT-VLAD	4608
VZClassifier	MR8filters	VZ-MR8	4096
Wavelet Statistical and Co-occurrence Features	Haar wavelet	WSF + WCF _{haar}	84
Wavelet Statistical and Co-occurrence Features	Bi-orthogonal wavelet	WSF + WCF _{bior22}	84
<i>Colour texture descriptors</i>			
Improved Opponent Colour LBP	Rotation-invariant	IOCLBP	1287
Integrative Co-occurrence Matrices		ICM	360
Integrative Co-occurrence Matrices	Rotation-invariant	ICM ^{DFT}	360
Local Binary Patterns + Local Colour Contrast	Rotation-invariant LBP	LBP + LCC	876
Local Colour Vector Binary Patterns	Rotation-invariant	LCVBP	432
Opponent Colour Local Binary Patterns	Rotation-invariant	OCLBP	648
Opponent Gabor features		OppGabor	630
Opponent Gabor features	Contrast-normalised	OppGabor _{cn}	630
Opponent Gabor features	Rotation-onvariant	OppGabor ^{DFT}	630
Opponent Gabor features	Rot.-inv. and ctr.-norm.	OppGabor _{cn} ^{DFT}	630

Table 3. Summary table of the off-the-shelf CNN-based features used in the experiments.

Pre-Trained Model	Output Layer (No. or Name)	Aggregation Method	Abbreviation	No. of Features
DenseNet_161.caffemodel	“pool5” (last Fully-Conn.)	None	DenseNet-161-FC	2208
DenseNet_161.caffemodel	“concat_5_24” (last conv.)	BoVW	DenseNet-161-BoVW	2208
DenseNet_201.caffemodel	“pool5” (last fully-conn.)	None	DenseNet-201-FC	1920
DenseNet_201.caffemodel	“concat_5_32” (last conv.)	BoVW	DenseNet-201-BoVW	1920
imagenet-googlenet-dag	“cls3_pool” (last fully-conn.)	None	GoogLeNet-FC	1024
imagenet-googlenet-dag	“icp9_out” (conv.)	BoVW	GoogLeNet-BoVW	1024
imagenet-caffe-alex	20 (last fully-conn.)	None	Caffe-Alex-FC	4096
imagenet-caffe-alex	13 (last conv.)	BoVW	Caffe-Alex-BoVW	4096
imagenet-caffe-alex	13 (last conv.)	VLAD	Caffe-Alex-VLAD	4224
imagenet-resnet-50-dag	“pool5” (last fully-conn.)	None	ResNet-50-FC	2048
imagenet-resnet-50-dag	“res5c_branch2c” (conv.)	BoVW	ResNet-50-BoVW	2048
imagenet-resnet-101-dag	“pool5” (last fully-conn.)	None	ResNet-101-FC	2048
imagenet-resnet-101-dag	“res5c_branch2c” (conv.)	BoVW	ResNet-101-BoVW	2048
imagenet-resnet-152-dag	“pool5” (last fully-conn.)	None	ResNet-152-FC	2048
imagenet-resnet-152-dag	“res5c_branch2c” (conv.)	BoVW	ResNet-152-BoVW	2048
imagenet-vgg-f	20 (last fully-conn.)	None	VGG-F-FC	4096
imagenet-vgg-f	13 (last conv.)	BoVW	VGG-F-BoVW	4096
imagenet-vgg-f	13 (last conv.)	VLAD	VGG-F-VLAD	4096
imagenet-vgg-m	20 (last fully-conn.)	None	VGG-M-FC	4096
imagenet-vgg-m	13 (last conv.)	BoVW	VGG-M-BoVW	4096
imagenet-vgg-m	13 (last conv.)	VLAD	VGG-M-VLAD	4096
imagenet-vgg-s	20 (last fully-conn.)	None	VGG-S-FC	4096
imagenet-vgg-s	13 (last conv.)	BoVW	VGG-S-BoVW	4096
imagenet-vgg-s	13 (last conv.)	VLAD	VGG-S-VLAD	4096
imagenet-vgg-verydeep-16	36 (last fully-conn.)	None	VGG-VD-16-FC	4096
imagenet-vgg-verydeep-16	29 (last conv.)	BoVW	VGG-VD-16-BoVW	4096
imagenet-vgg-verydeep-16	29 (last conv.)	VLAD	VGG-VD-16-VLAD	4096
imagenet-vgg-verydeep-19	42 (last fully-conn.)	None	VGG-VD-19-FC	4096
imagenet-vgg-verydeep-19	35 (last conv.)	BoVW	VGG-VD-19-BoVW	4096
imagenet-vgg-verydeep-19	35 (last conv.)	VLAD	VGG-VD-19-VLAD	4096
vgg-face	36 (last fully-conn.)	None	VGG-Face-FC	4096
vgg-face	29 (last conv.)	BoVW	VGG-Face-BoVW	4096
vgg-face	29 (last conv.)	VLAD	VGG-Face-VLAD	4096

4.1. Hand-Designed Descriptors

4.1.1. Purely Spectral Descriptors

- Mean (*Mean*): Average of each of the R, G and B channels (three features).
- Mean + standard deviation (*Mean + Std*): Average and standard deviation of each of the R, G and B channels (six features).
- Mean + moments (*Mean + Moms*): Average and central moments from second to second of each of the R, G and B channels (15 features) [64].
- Quartiles (*Quartiles*): The 25%, 50% and 75% percentile of each colour channel (nine features) [64].
- Marginal histograms (*Marg-Hists-256*): Concatenation of the 256-bin marginal histograms of the R, G and B channels ($3 \times 256 = 768$ features) [57,65].
- Joint colour histogram (*Full-Hist-10*): Full 3D colour histogram in the RGB space [66] with 10 bins per channel ($10^3 = 1000$ features).

4.1.2. Grey-Scale Texture Descriptors

Histograms of Equivalent Patterns

Six LBP variants, also referred to as histograms of equivalent patterns [67], specifically:

- *Completed Local Binary Patterns* [68] (CLBP, 324 features)

- Gradient-Based Local Binary Patterns [69] (GLBP, 108 features)
- Improved Local Binary Patterns [70] (ILBP, 213 features)
- Local Binary Patterns [71] (LBP, 108 features)
- Local Ternary Patterns [72] (LTP, 216 features)
- Texture Spectrum [73] (TS, 2502 features)

Each of the above methods was used to obtain a multiple resolution feature vector by concatenating the rotation-invariant vectors (e.g., LBP^{ri}) computed at resolutions of 1 px, 2 px and 3 px [74]. For each resolution, we used non-interpolated neighbourhoods [75] composed of a central pixel and eight peripheral pixels, as shown in Figure 11.

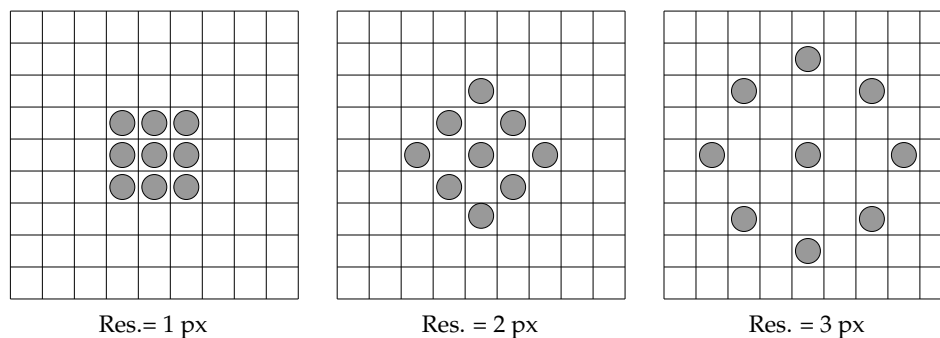


Figure 11. Neighbourhoods used to compute features through histograms of equivalent patterns and other methods.

- Grey-Level Co-occurrence Matrices (GLCM): Five global statistics (i.e., *contrast*, *correlation*, *energy*, *entropy* and *homogeneity*) from grey-level co-occurrence matrices [76] computed using displacement vectors of lengths of 1 px, 2 px and 3 px and orientations of 0° , 45° , 90° and 135° ($5 \times 3 \times 4 = 60$ features). A rotation-invariant version ($GLCM^{DFT}$) based on discrete Fourier transform normalisation [77] was also considered.
- Histograms of Oriented Gradients (HOG): Concatenation of three densely-computed 256-dimensional histograms of oriented gradients [78] ($3 \times 256 = 768$ features) estimated through Sobel filters of dimensions of $3 \text{ px} \times 3 \text{ px}$, $5 \text{ px} \times 5 \text{ px}$ and $7 \text{ px} \times 7 \text{ px}$.
- Image Patch-Based Classifier, Joint version (IPBC-J): Local image patches aggregated over a dictionary of visual words (Section 4.3) as proposed by Varma and Zissermann [79]. The image patches were captured at resolutions of 1 px, 2 px and 3 px using the same neighbourhood configuration shown in Figure 11. The resulting feature vectors were concatenated into a single vector. Further pre-processing involved zero-mean and unit-variance normalisation of the input image and contrast normalisation of each patch through Weber's law, as recommended in [79].
- Gabor features (Gabor): Mean and standard deviation of the magnitude of Gabor-transformed images from a bank of filters with five frequencies and seven orientations. The other parameters of the filter bank were: frequency spacing half octave, spatial-frequency bandwidth one octave and aspect ratio 1.0. We considered both raw and contrast-normalised response: in the second case, the magnitudes for one point in all frequencies and rotations were normalized to sum to one. This option is indicated with subscript "cn" in the remainder. For both options, a DFT-based rotation-invariant version [80] (superscript "DFT" in the remainder) was also included in the experiments. In all the versions, the number of features was $2 \times 5 \times 7 = 70$.
- Wavelets (WSF + WCF): Statistical (mean and standard deviation) and Co-occurrence features (same as in GLCM) from a three-level Wavelet decomposition as described in [81]. We used Haar's and bi-orthogonal wavelets, respectively indicated with subscript "haar" and "bior22" in the remainder. The number of statistical features was $2 \times 4 \times 3 = 24$, and that of the co-occurrence features was $6 \times 4 \times 3 = 60$, making a total of 84 features.

- VZ classifier with MR8 filters (*VZ-MR8*): Filter responses from a bank of 36 anisotropic filters (first- and second-derivative filters at six orientations and three scales) plus two rotationally-symmetric ones (a Gaussian and a Laplacian of Gaussian). Only eight responses are retained, i.e., the six maximum responses at each scale across all orientations and the response of the anisotropic filters [82]. The filter responses were aggregated over a dictionary of 4096 visual words (see Section 4.3).
- Dense SIFT (*SIFT-BoVW*): Spatial histograms of local gradient orientations computed every two pixels and over a neighbourhood of radius 3 px (histograms of equivalent patterns). The resulting 128-dimensional local features were aggregated over a dictionary of 4096 visual words as described in Section 4.3.
- Dense SIFT (*SIFT-VLAD*): Same settings as SIFT-BoVW, but with Vectors of Locally-Aggregated Descriptors (VLAD) aggregation (Section 4.3) over 32 clusters.

4.1.3. Colour Texture Descriptors

- Integrative Co-occurrence Matrices (*ICM*): Co-occurrence features extracted from each colour channel separately and from the R-G, R-B and G-B pairs of channels as described in [83,84]. The other settings (i.e., statistics and displacement vectors) were the same as in GLCM (Section 4.1.2). The feature vector is six-times longer than GLCM's, resulting in $60 \times 6 = 360$ features. A rotation-invariant version (ICM^{DFT}) based on the same scheme as $GLCM^{DFT}$ was also considered.
- Local Binary Patterns + Local Colour Contrast (*LBP + LCC*): Concatenation of Local Binary Patterns and Local Colour Contrast (LCC) features as described in [85]. LCC is the probability distribution (histogram) of the angle between the colour vector of the central pixel in the neighbourhood and the average colour vector of the peripheral pixels. Following the settings suggested in [85], we used histograms of 256 bins for each resolution (i.e., 1 px, 2 px and 3 px) and concatenated the results. Concatenation with LBP gives a total of $108 + 256 \times 3 = 876$ features.
- Local Colour Vector Local Binary Patterns (*LCVBP*): Concatenation of Colour Norm Patterns (CNP) and Colour Angular Patterns (CAP) as proposed by Lee et al. [86]. In CNP, the colour norm of a pixel in the periphery is thresholded at the value of the central pixel; in CAP, the thresholding is based on the angle that the projections of the colour vectors form in the RG, RB and GB planes. Since the CNP feature vector is the same length as LBP and CAP's is three-times longer, their concatenation produces $108 \times 4 = 432$ features.
- Opponent Colour Local Binary Patterns (*OCLBP*): Local Binary Patterns computed on each colour channel separately and from the R-G, R-B and G-B pairs of channels [87]. The other settings (type of neighbourhood and features) were the same as in grey-scale LBP (Section 4.1.2). The resulting feature vector is six-times longer than LBP's; therefore, we have $108 \times 6 = 648$ features.
- Improved Opponent Colour Local Binary Patterns (*IOCLBP*): An improved version of OCLBP in which thresholding is point-to-average instead of point-to-point [88]. This can be considered a colour version of ILBP (see Section 4.1.2).
- Opponent Gabor Features (*OGF*): A multi-scale representation including intra- and inter-channel features as proposed in [89]. This comprises 2×3 (channels) $\times 5$ (frequencies) $\times 7$ (orientations) = 210 monochrome features from each colour channel and 2×3 (channels) $\times 10$ (combinations of frequencies) $\times 7 = 420$ intra-channel features. The total number of features is $210 + 420 = 630$.

4.2. Off-the-Shelf CNN-Based Features

The CNN-based features were computed using the following 13 pre-trained models:

- Caffe-Alex [4]
- DenseNet-161 and DenseNet-201 [90]
- GoogLeNet [91]

- ResNet-50, ResNet-101 and ResNet-152 [92]
- VGG-F, VGG-M and VGG-S [93]
- VGG-VD-16 and VGG-VD-19 [6]
- VGG-Face [7]

Twelve of the above models were trained for object recognition and the remaining one (vgg-face) for face recognition. Each network was used as a generic feature extractor, and the resulting features were passed on to a standard classifier (see Section 5). Following the strategy suggested in recent works [16,17,57], we considered the following two alternative types of features (see Section 4.3 and Figure 12):

- The order-sensitive output of the last fully-connected layer;
- The aggregated, orderless output of the last convolutional layer.

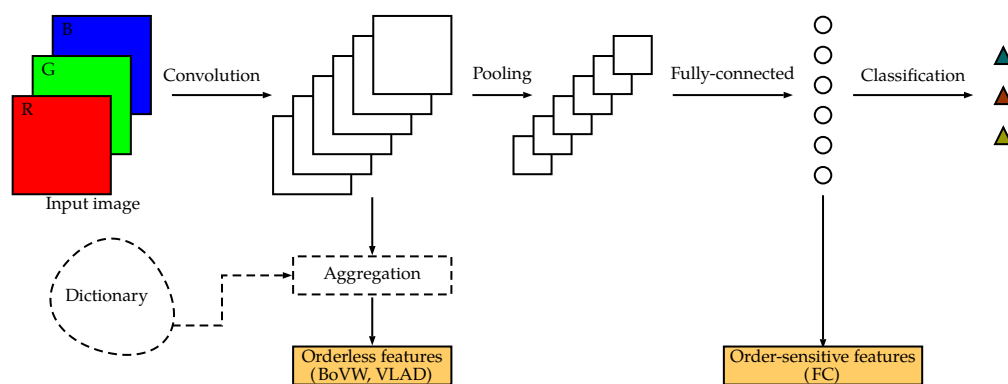


Figure 12. Simplified diagram of a generic convolutional neural network. For texture classification, we can either use the order-sensitive output of a fully-connected layer or the orderless, aggregated output of a convolutional layer.

4.3. Learned vs. Unlearned Methods: Aggregation

All the methods considered in the experiments (and, more generally, all local image descriptors) can be either *learned* or *unlearned*. Methods belonging to the first group are also referred to as *a posteriori* and those belonging to the second as *a priori* (for a discussion on this point, see also [67,94]). In the first group, the image-to-features mapping is the result of a preliminary learning stage aiming at generating a dictionary of visual words upon which the local features are aggregated. By contrast, unlearned methods do not require any training phase, since the image-to-feature mapping is a priori and universal.

Aggregation

Aggregation (also referred to as *pooling* [16]) is the process by which local image features (for instance, the output of a bank of filters or that of a layer of a convolutional network) is grouped around a dictionary of visual words in order to obtain a global feature vector suitable for being used with a classifier [95].

The first step of the aggregation process is the definition of the dictionary, which usually consists of vector-quantizing the local features into a set of prototypes. Key factors in this phase are:

- The dimension of the dictionary;
- The algorithm used for clustering;
- The set of images used for training.

In our experiments, the dimension of the dictionary depended on the feature encoder used, as discussed below. The algorithm for clustering was always the *k-means*; whereas for training, we followed the same approach as in [57]; i.e., we used a set of colour texture images from an external source [96]. To avoid overfitting and possibly biased results, these images were different from those contained in any of the datasets detailed in Section 3.

For the aggregation, we considered two schemes (due to dimensionality reasons, aggregation over the DenseNet, GoogLeNet and ResNet models was limited to BoVW) [16,95]:

- Bag of Visual Words (BoVW);
- Vectors of Locally-Aggregated Descriptors (VLAD).

This choice was based on recent works [16,57] and was also the result of a trade-off between accuracy and dimensionality (recall that for a D -dimensional feature space and a dictionary with K visual words, the number of features respectively generated by BoVW and VLAD is K and $K \times D$).

Convolutional networks also have built-in aggregation modules: the Fully-Connected (FC) layers. However, whereas BoVW and VLAD implement orderless aggregation (i.e., they discard the spatial configuration of the features), the aggregation provided by fully-connected layers is order-sensitive. The number of features produced by FC layers depends on the network's architecture and is therefore fixed. For a fair comparison between the three aggregation strategies (FC, BoVW and VLAD), we chose a number of visual words for BoVW and VLAD that produced a number of features as close as possible to that produced by FC.

Post-processing involved L_1 normalization of the BoVW features and L_2 normalization of the individual VLAD vectors and vectors of FC features [16,57].

5. Experiments

We comparatively evaluated the discrimination accuracy and computational demand of the methods detailed in Section 4 on a set of supervised image classification tests using the datasets described in Section 3. In the remainder 'Experiment # N ' will indicate the experiment ran on the colour texture images of Group # N . Following the same approach as in recent related works [1,14,17,18], we used the nearest-neighbour classification strategy (with L_1 distance) in all the experiments.

Performance evaluation was based on split-sample validation with stratified sampling. The fraction of samples of each class used to train the classifier was 1/2 for Experiments #1 and #2 and 1/8 for all the other experiments. In the first two cases, the choice was dictated by the low number of samples available (as few as four per class in some datasets); in the others, we opted for a lower training ratio in order to better estimate the robustness of the methods to the intra-class variability. The figure of merit ('accuracy' in the remainder) was the ratio between the number of samples of the test set classified correctly and the total number of samples of the test set. For a stable estimation, the value was averaged over 100 different subdivisions into a training and test set.

For each experiment, a ranking of method was obtained by comparing all the image descriptors pairwise and respectively assigning +1, -1 or 0 each time a method was significantly better, worse or not significantly different from the other. Statistical significance ($\alpha = 0.05$) was assessed through the Wilcoxon–Mann–Whitney rank sum test [97] over the accuracy values resulting from the 100 splits.

6. Results and Discussion

6.1. Accuracy

Tables 4 and 5 respectively report the relative performance in terms of ranking (as defined in Section 5) of the ten best hand-crafted descriptors and ten best CNN-based features. The results depict a scenario, which, on the whole, was dominated by CNN-based methods. Among them, the three ResNet outperformed by far the other networks, and interestingly, the FC configuration emerged as the best strategy to extract CNN-based features among the three considered (FC, BoVW and VLAD).

Conversely, the hand-crafted descriptors came lower in the standings and were dominated by colour variants of LBP (e.g., IOCLBP, OCLBP and LCVBP).

Table 4. Hand-crafted descriptors: relative performance of the best ten methods at a glance. For each method, the columns from #1–#10 show the rank (first row) and average accuracy (second row, in parentheses) by experiment. The next to last column reports the average rank and accuracy over all the experiments and the last column the overall position in the placings.

Descriptor	Rank (by Experiment)										Avg.	Overall Position
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10		
IOCLBP ⁽⁹⁾	64.0 (91.9)	55.0 (78.6)	54.0 (84.4)	52.0 (71.6)	64.5 (96.8)	63.0 (82.4)	48.5 (74.0)	66.0 (71.7)	49.0 (67.9)	30.0 (18.6)	54.6 (73.8)	9
OCLBP ⁽¹⁴⁾	63.0 (91.5)	54.0 (77.3)	50.0 (81.8)	34.5 (68.1)	68.0 (97.2)	61.0 (80.3)	44.0 (72.0)	64.5 (70.2)	44.0 (66.3)	20.0 (16.7)	50.3 (72.1)	14
ICM ^{DFT}	61.0 (90.6)	47.0 (76.9)	30.5 (73.5)	25.0 (61.4)	64.5 (96.5)	57.5 (76.4)	51.0 (74.9)	61.0 (68.0)	38.0 (63.5)	21.0 (16.9)	45.6 (69.9)	17
LCVBP	60.0 (91.1)	59.5 (82.1)	42.0 (77.4)	53.5 (70.0)	53.0 (93.6)	59.0 (77.8)	12.5 (57.1)	42.5 (60.6)	43.0 (66.1)	23.0 (17.1)	44.8 (69.3)	18
Full-Hist-10	46.0 (83.1)	53.0 (75.8)	11.0 (56.2)	26.0 (64.4)	45.0 (92.5)	67.0 (84.4)	61.5 (81.1)	67.5 (73.3)	40.0 (63.3)	11.0 (14.4)	42.8 (68.8)	21
ILBP	49.0 (86.8)	35.0 (70.6)	51.0 (81.5)	60.0 (69.9)	52.0 (94.2)	47.5 (71.6)	22.0 (60.9)	45.5 (61.4)	26.0 (59.5)	36.0 (21.0)	42.4 (67.7)	22
ICM	55.0 (90.1)	43.0 (75.9)	21.0 (71.2)	18.0 (59.0)	61.0 (95.0)	53.5 (73.4)	50.0 (73.6)	60.0 (67.4)	37.0 (63.0)	12.5 (15.0)	41.1 (68.4)	23
LBP + LCC	56.5 (90.5)	42.0 (74.5)	43.0 (78.4)	58.5 (69.7)	49.0 (94.0)	53.5 (72.9)	18.0 (58.4)	40.0 (59.3)	29.0 (59.9)	17.5 (16.6)	40.7 (67.4)	25
SIFT-BoVW	38.0 (84.5)	44.5 (76.4)	60.0 (87.1)	62.0 (72.4)	19.0 (86.1)	20.5 (59.4)	27.0 (64.1)	53.0 (64.4)	31.0 (61.6)	46.0 (31.6)	40.1 (68.8)	27
CLBP	47.0 (87.8)	38.0 (72.4)	36.0 (75.5)	53.5 (66.9)	47.0 (93.6)	51.5 (72.5)	21.0 (59.3)	37.5 (58.1)	32.0 (59.9)	35.0 (20.6)	39.9 (66.7)	28

Table 5. CNN-based descriptors: relative performance of the best ten methods at a glance. For each method, the columns from #1–#10 show the rank (first row) and average accuracy (second row, in parentheses) by experiment. The next to last column reports the average rank and accuracy over all the experiments and the last column the overall position in the placings.

Descriptor	Rank (by Experiment)										Avg.	Overall Position
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10		
ResNet-50-FC	68.0 (91.4)	68.0 (87.9)	68.0 (94.8)	68.0 (84.8)	66.0 (96.1)	68.0 (84.6)	66.0 (84.5)	64.5 (70.4)	66.0 (83.3)	68.0 (60.8)	67.0 (83.9)	1
ResNet-101-FC	66.0 (90.5)	67.0 (87.1)	67.0 (94.4)	67.0 (84.0)	67.0 (95.9)	65.5 (83.9)	68.0 (86.0)	62.0 (69.2)	67.5 (83.4)	66.5 (60.3)	66.3 (83.5)	2
ResNet-152-FC	65.0 (90.4)	66.0 (87.0)	66.0 (94.1)	66.0 (83.7)	62.0 (95.7)	65.5 (83.9)	67.0 (85.5)	63.0 (69.5)	67.5 (83.6)	66.5 (60.4)	65.5 (83.4)	3
VGG-VD-16-FC	54.0 (88.4)	64.0 (83.3)	64.0 (91.6)	64.0 (81.3)	58.0 (94.1)	62.0 (82.2)	64.0 (83.2)	59.0 (66.4)	64.0 (79.6)	64.5 (56.2)	61.8 (80.6)	4
VGG-VD-19-FC	50.0 (87.1)	59.5 (82.3)	65.0 (91.7)	65.0 (81.3)	51.0 (93.4)	60.0 (81.6)	65.0 (83.6)	57.5 (65.8)	65.0 (79.8)	64.5 (56.2)	60.3 (80.3)	5
VGG-M-FC	59.0 (88.6)	65.0 (82.5)	63.0 (91.3)	63.0 (79.6)	54.0 (93.6)	57.5 (79.4)	63.0 (79.3)	54.5 (64.7)	63.0 (78.3)	59.0 (50.0)	60.1 (78.7)	6
VGG-S-FC	62.0 (89.3)	62.5 (81.8)	62.0 (90.9)	61.0 (78.9)	59.0 (94.0)	56.0 (79.3)	58.5 (78.1)	54.5 (64.8)	62.0 (78.2)	62.5 (51.3)	60.0 (78.7)	7
VGG-F-FC	58.0 (88.7)	61.0 (81.7)	61.0 (90.4)	58.5 (78.1)	56.5 (93.8)	51.5 (77.4)	57.0 (77.9)	56.0 (65.2)	60.5 (77.3)	58.0 (46.8)	57.8 (77.7)	8
Caffe-Alex-FC	53.0 (88.4)	51.5 (77.5)	59.0 (89.2)	56.5 (76.1)	56.5 (94.0)	42.0 (73.3)	60.0 (78.6)	50.0 (62.8)	60.5 (76.6)	56.0 (43.6)	54.5 (76.0)	10
VGG-S-VLAD	52.0 (87.7)	62.5 (81.8)	57.0 (84.7)	50.5 (74.1)	60.0 (94.0)	50.0 (76.8)	53.0 (75.2)	45.5 (61.3)	57.0 (75.9)	57.0 (46.3)	54.5 (75.8)	11

It is also most informative to analyse the results by experiment and dataset as reported in Tables 6–16. For each experiment, the corresponding table shows the average accuracy attained by the

best five hand-crafted image descriptors and the best five CNN-based features (the other values are provided as Supplementary Material).

Table 6. Results of Experiment #1 (Part 1: Datasets 1–7): stationary textures acquired under steady imaging conditions. Figures report overall accuracy by dataset. Boldface denotes the highest value; underline signals a statistically-significant difference between the best hand-crafted and the best CNN-based descriptor. Descriptors are listed in ascending order of by-experiment rank (best first).

Descriptor	Dataset							Avg.
	1	2	3	4	5	6	7	
<i>Hand-crafted</i>								
IOCLBP	96.7	99.5	98.7	100.0	83.6	98.4	<u>95.3</u>	91.9
OCLBP	96.0	99.5	96.6	100.0	85.3	99.1	95.2	91.5
ICM ^{DFT}	95.8	99.0	94.2	100.0	95.7	99.2	91.7	90.6
LCVBP	92.4	97.7	85.7	100.0	<u>97.4</u>	98.0	88.2	91.1
LBP + LCC	94.4	97.3	89.7	99.7	90.5	98.1	87.4	90.5
<i>CNN-based</i>								
ResNet-50-FC	98.7	100.0	87.8	100.0	96.6	99.0	90.6	91.4
DenseNet-161-FC	98.6	99.7	<u>100.0</u>	100.0	83.2	99.5	90.6	91.5
ResNet-101-FC	98.9	99.8	85.7	100.0	95.5	99.3	89.9	90.5
ResNet-152-FC	98.7	100.0	84.1	100.0	93.4	98.8	89.6	90.4
VGG-S-FC	98.6	99.3	87.4	100.0	88.3	98.7	86.4	89.3

Datasets: 1) ALOT-95-S-N; 2) CBT-99-S-N; 3) Drexel-18-S-N; 4) KylbergSintorn-25-S-N; 5) MBT-120-S-N; 6) MondialMarmi-25-S-N; 7) Outex-192-S-N.

Table 7. Results of Experiment #1 (Part 2: Datasets 8–14): stationary textures acquired under steady imaging conditions. Figures report overall accuracy by dataset. Boldface denotes the highest value; underline signals a statistically-significant difference between the best hand-crafted and the best CNN-based descriptor. Descriptors are listed in ascending order of by-experiment rank (best first).

Descriptor	Dataset							Avg.
	1	2	3	4	5	6	7	
<i>Hand-crafted</i>								
IOCLBP	69.0	77.7	99.2	97.3	94.2	82.1	<u>95.4</u>	91.9
OCLBP	69.7	76.9	98.8	96.2	92.3	83.0	93.8	91.5
ICM ^{DFT}	62.1	76.8	98.0	93.9	92.5	81.5	90.1	90.6
LCVBP	75.1	75.4	98.4	96.6	95.2	83.7	91.5	91.1
LBP + LCC	67.2	79.3	99.1	93.7	92.6	85.7	94.2	90.5
<i>CNN-based</i>								
ResNet-50-FC	53.0	86.6	99.8	99.0	99.6	85.4	85.4	91.4
DenseNet-161-FC	69.0	75.1	99.7	97.5	96.2	82.9	88.5	91.5
ResNet-101-FC	51.8	82.1	99.7	98.8	99.2	85.1	83.8	90.5
ResNet-152-FC	56.1	83.9	99.7	98.6	99.4	83.2	82.9	90.4
VGG-S-FC	56.3	74.3	98.1	98.1	98.0	86.0	82.3	89.3

Datasets: 1) ALOT-95-S-N; 2) CBT-99-S-N; 3) Drexel-18-S-N; 4) KylbergSintorn-25-S-N; 5) MBT-120-S-N; 6) MondialMarmi-25-S-N; 7) Outex-192-S-N; 8) Parquet-38-S-N; 9) PlantLeaves-20-S-N; 10) RawFoot-68-S-N; 11) STex-202-S-N; 12) USPTex-137-S-N; 13) VisTex-89-S-N; 14) V×CTSG-42-S-N.

Experiments #1 and #2 (Tables 6–8) show that, under steady imaging conditions, hand-crafted-descriptors were competitive only with stationary textures, whereas CNN-based features proved clearly superior with non-stationary ones (Figures 13–15). In Experiment #1 (Tables 6 and 7), the best-performing method belonged to the first group in the four datasets out of 14; the reverse occurred in eight datasets, whereas in the remaining two, the difference did not reach statistical significance (Figure 16). Interestingly, there was a marked gap when it came to classifying fine textures

with a high degree of similarity, such as in datasets Parquet-38-S-N and V×CTSG-42-S-N (Figure 14). In this case, the hand-crafted descriptors outperformed CNN-based features by a good margin. With non-stationary textures (Experiment #2), CNN-based features proved generally better, outperforming hand-crafted descriptors in six datasets out of nine, whereas the reverse occurred in two datasets only.

Table 8. Results of Experiment #2: non-stationary textures acquired under steady imaging conditions. Figures report overall accuracy by dataset. Boldface denotes the highest value; underline signals a statistically-significant difference between the best hand-crafted and the best CNN-based descriptor. Descriptors are listed in ascending order of by-experiment rank (best first).

Descriptor	Dataset								Avg.
	1	2	3	4	5	6	7	8	
<i>Hand-crafted image descriptors</i>									
LCVBP	75.4	75.3	89.8	85.2	77.6	88.6	95.8	69.1	82.1
IOCLBP	78.6	79.9	54.4	80.2	84.1	90.3	95.4	66.0	78.6
OCLBP	74.8	74.8	60.1	79.9	83.3	87.7	92.6	65.0	77.3
Full-Hist-10	87.7	87.7	28.8	78.0	79.3	87.1	92.2	65.6	75.8
Marginal-Hists-256	71.0	79.6	99.6	73.7	81.5	70.5	79.6	70.3	78.2
<i>CNN-based features</i>									
ResNet-50-FC	94.9	93.0	77.7	90.7	78.5	97.9	99.6	71.0	87.9
ResNet-101-FC	94.7	92.8	76.6	90.5	76.6	97.5	99.2	68.6	87.1
ResNet-152-FC	95.0	91.8	76.5	90.9	76.7	97.4	99.1	68.8	87.0
VGG-M-FC	87.7	86.8	64.1	84.3	74.5	95.1	99.2	68.3	82.5
VGG-VD-16-FC	93.1	85.6	71.5	79.1	76.8	96.3	96.6	67.2	83.3

Datasets: 1) ALOT-40-NS-N; 2) ForestSpecies-112-NS-N; 3) MBT-34-NS-N; 4) NewBarkTex-6-NS-N; 5) Outex-59-NS-N; 6) STex-138-NS-N; 7) USPTex-33-NS-N; 8) VisTex-78-NS-N.

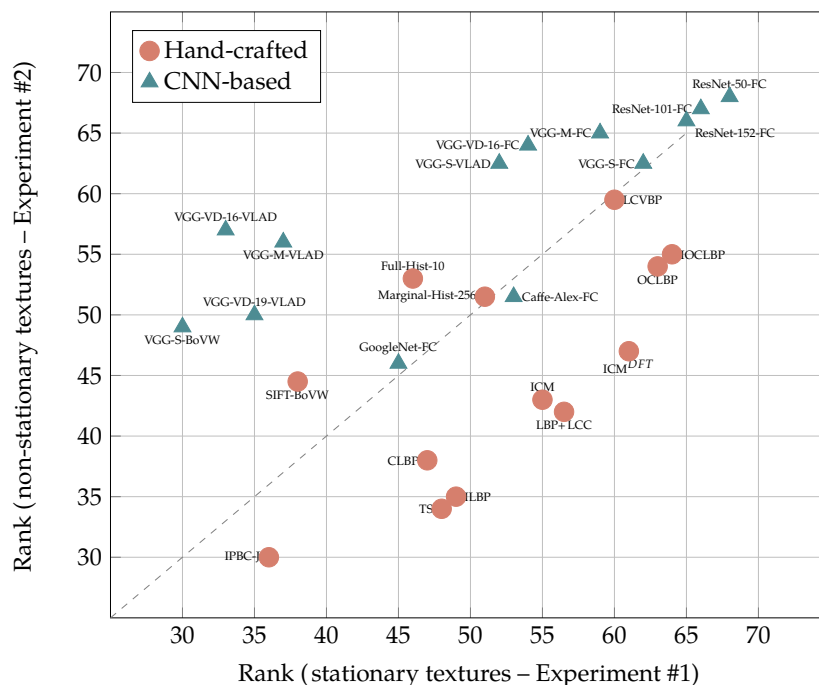


Figure 13. Relative performance of hand-crafted descriptors vs. CNN-based features with stationary (x axis) and non-stationary textures (y axis) under invariable imaging conditions (Experiments #1 and #2, best 13 methods). The plot shows a clear divide between CNN-based methods (mostly clustered in the upper-left part, therefore showing affinity for non-stationary textures) and hand-crafted descriptors (mostly clustered in the lower-right part, therefore showing affinity for stationary textures)

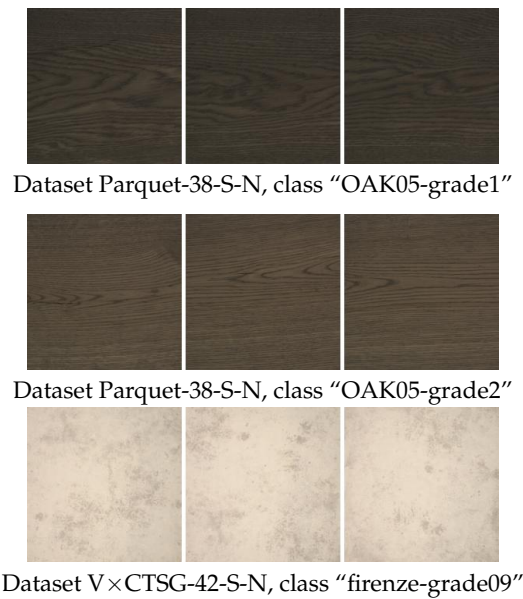


Figure 14. Hand-crafted image descriptors proved generally better than CNN-based features for the classification of stationary textures acquired under invariable imaging conditions. The gap was substantial when it came to discriminating between very similar textures, as those shown in the picture.

Under variable illumination conditions, CNN-based descriptors seemed to be able to compensate for changes in illumination better than hand-crafted descriptors did (Experiments #3 and #4, Tables 9–10). This result is in agreement with the findings of Cusano et al. [57].

Table 9. Results of Experiment #3: stationary textures with variations in illumination. Figures report overall accuracy by dataset. Boldface denotes the highest value; underline signals a statistically-significant difference between the best hand-crafted and the best CNN-based descriptor. Descriptors are listed in ascending order of by-experiment rank (best first).

Descriptor	Dataset					Avg.
	1	2	3	4	5	
<i>Hand-crafted image descriptors</i>						
SIFT-BoVW	77.9	79.0	95.1	97.5	86.2	87.1
IOCLBP	82.0	85.9	91.7	82.8	79.5	84.4
IPBC-J	70.5	85.7	89.6	94.2	71.6	82.3
VZ-MR8	73.3	69.8	94.1	96.4	82.0	83.1
ILBP	76.9	84.2	80.8	96.4	69.0	81.5
<i>CNN-based features</i>						
ResNet-50-FC	96.0	83.0	99.1	98.8	97.2	94.8
ResNet-101-FC	95.6	81.6	98.9	98.9	97.0	94.4
ResNet-152-FC	95.5	81.9	98.8	98.3	95.9	94.1
VGG-VD-19-FC	93.4	76.4	96.8	97.5	94.4	91.7
VGG-VD-16-FC	93.9	76.1	97.2	97.2	93.9	91.6

Datasets: 1) ALOT-95-S-I; 2) Outex-192-S-I; 3) RawFooT-68-S-I-1; 4) RawFooT-68-S-I-2; 5) RawFooT-68-S-I-3.

The results were however rather split in the presence of rotation (Experiments #5 and #6, Tables 11 and 12). Here, the hand-designed descriptors were significantly better than CNN-based features in three datasets out of six, whereas the reverse occurred in two datasets. This parallels the results reported in [1] and is likely to be related to the directional nature of the learned filters in the convolutional networks.

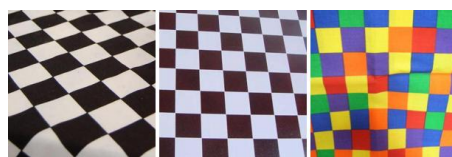
A similar trend emerged with variations in scale (Experiments #7 and #8, Tables 13 and 14). In this case, the hand-designed descriptors (particularly 3D colour histogram) were significantly better than CNN-based features in two datasets, while the reverse occurred in the other two.

Table 10. Results of Experiment #4: non-stationary textures with variations in illumination. Figures report overall accuracy by dataset. Boldface denotes the highest value; underline signals a statistically-significant difference between the best hand-crafted and the best CNN-based descriptor. Descriptors are listed in ascending order of by-experiment rank (best first).

Descriptor	Dataset		Avg.
	1	2	
<i>Hand-crafted image descriptors</i>			
SIFT-BoVW	70.1	74.7	72.4
ILBP	63.1	<u>76.7</u>	69.9
LBP + LCC	64.6	74.9	69.7
CLBP	59.7	74.0	66.9
LCVBP	69.2	70.9	70.0
<i>CNN-based features</i>			
ResNet-50-FC	<u>95.7</u>	74.0	84.8
ResNet-101-FC	<u>95.7</u>	72.3	84.0
ResNet-152-FC	<u>95.7</u>	71.7	83.7
VGG-VD-19-FC	93.4	69.2	81.3
VGG-VD-16-FC	93.9	68.7	81.3

Datasets: 1) ALOT-40-NS-I; 2) Outex-59-NS-I.

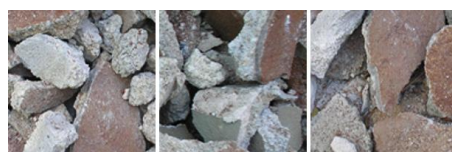
In the presence of multiple and uncontrolled changes in the imaging conditions, including variations in illumination, scale and viewpoint (Experiments #9 and #10, Tables 15 and 16), the hand-crafted descriptors were just non-competitive: CNN-based features proved superior in all the datasets considered. The difference was more noticeable in those datasets (e.g., RDAD) where the intra-class variability was higher. Particularly interesting were the results obtained with the Describable Texture Dataset: here, CNN-based features surpassed hand-crafted descriptors by ≈ 30 percentage points. On the same dataset, the 60.8% accuracy achieved by ResNet-50 was equally remarkable, in absolute terms.



Dataset DTD-47-NS-M, class "Chequered"



Dataset DTD-47-NS-M, class "Cracked"



Dataset USPTex-33-NS-N, class "c108"

Figure 15. CNN-based features were on the whole better than hand-crafted descriptors at classifying non-stationary textures, as those shown in the picture.

Table 11. Results of Experiment #5: stationary textures with rotation. Figures report overall accuracy by dataset. Boldface denotes the highest value; underline signals a statistically-significant difference between the best hand-crafted and the best CNN-based descriptor. Descriptors are listed in ascending order of by-experiment rank (best first).

Descriptor	Dataset				Avg.
	1	2	3	4	
<i>Hand-crafted image descriptors</i>					
OCLBP	94.9	100.0	99.2	<u>94.6</u>	97.2
IOCLBP	94.9	100.0	98.6	93.8	96.8
ICM ^{DFT}	95.6	100.0	<u>99.4</u>	91.2	96.5
ICM	93.6	100.0	99.0	87.6	95.0
LCVBP	90.4	99.9	98.1	86.0	93.6
<i>CNN-based features</i>					
ResNet-101-FC	98.4	100.0	98.3	86.9	95.9
ResNet-50-FC	<u>98.7</u>	100.0	98.0	87.6	96.1
DenseNet-161-FC	97.0	100.0	99.1	89.0	96.3
VGG-S-VLAD	96.8	100.0	98.3	81.0	94.0
VGG-S-FC	97.4	100.0	97.4	81.3	94.0

Datasets: 1) ALOT-95-S-R; 2) KylbergSintorn-25-S-R; 3) MondialMarmi-25-S-R; 4) Outex-193-S-R.

Table 12. Results of Experiment #6: non-stationary textures with rotation. Figures report overall accuracy by dataset. Boldface denotes the highest value; underline signals a statistically-significant difference between the best hand-crafted and the best CNN-based descriptor. Descriptors are listed in ascending order of by-experiment rank (best first).

Descriptor	Dataset		Avg.
	1	2	
<i>Hand-crafted image descriptors</i>			
Full-Hist-10	87.2	81.6	84.4
IOCLBP	81.2	<u>83.6</u>	82.4
OCLBP	77.4	83.3	80.3
LCVBP	79.2	76.4	77.8
ICM ^{DFT}	74.0	78.8	76.4
<i>CNN-based features</i>			
ResNet-50-FC	95.4	73.9	84.6
ResNet-101-FC	<u>95.5</u>	72.2	83.9
ResNet-152-FC	<u>95.5</u>	72.4	83.9
DenseNet-161-FC	86.0	76.4	81.2
VGG-VD-19-FC	93.8	69.4	81.6

Datasets: 1) ALOT-40-NS-R; 2) Outex-59-NS-R.

Table 13. Results of Experiment #7: stationary textures with variation in scale. Figures report overall accuracy by dataset. Boldface denotes the highest value; underline signals a statistically-significant difference between the best hand-crafted and the best CNN-based descriptor. Descriptors are listed in ascending order of by-experiment rank (best first).

Descriptor	Dataset			Avg.
	1	2	3	
<i>Hand-crafted image descriptors</i>				
Full-Hist-10	<u>89.1</u>	76.0	78.1	81.1
Marginal-Hist-256	76.0	70.9	81.7	76.2
Mean + Std	77.4	69.7	74.9	74.0
ICM ^{DFT}	65.2	79.3	80.2	74.9
ICM	64.0	77.7	79.1	73.6
<i>CNN-based features</i>				
ResNet-101-FC	84.3	<u>91.5</u>	82.1	86.0
ResNet-152-FC	85.7	89.0	81.9	85.5
ResNet-50-FC	81.7	88.9	<u>82.8</u>	84.5
VGG-VD-19-FC	86.8	87.3	76.7	83.6
VGG-VD-16-FC	87.1	86.2	76.2	83.2

Datasets: 1) KTH-TIPS-10-S-S; 2) KTH-TIPS-11b-S-S; 3) Outex-192-S-S.

Table 14. Results of Experiment #8: non-stationary textures with variation in scale. Figures report overall accuracy by dataset. Boldface denotes the highest value; underline signals a statistically-significant difference between the best hand-crafted and the best CNN-based descriptor. Descriptors are listed in ascending order of by-experiment rank (best first).

Descriptor	Dataset
	1
<i>Hand-crafted image descriptors</i>	
Full-Hist-10	73.3
Marginal-Hist-256	73.2
IOCLBP	71.7
OCLBP	70.2
ICM ^{DFT}	68.0
<i>CNN-based features</i>	
ResNet-50-FC	70.4
ResNet-101-FC	69.2
ResNet-152-FC	69.5
VGG-VD-16-FC	66.4
VGG-VD-19-FC	65.8

Datasets: 1) Outex-59-S-S.

Table 15. Results of Experiment #9: stationary textures with multiple variations. Figures report overall accuracy by dataset. Boldface denotes the highest value; underline signals a statistically-significant difference between the best hand-crafted and the best CNN-based descriptor. Descriptors are listed in ascending order of by-experiment rank (best first).

Descriptor	Dataset						Avg.
	1	2	3	4	5	6	
<i>Hand-crafted image descriptors</i>							
IOCLBP	88.7	33.8	84.1	91.7	66.9	41.9	67.9
OCLBP	89.1	30.9	82.2	90.3	64.4	41.0	66.3
LCVBP	90.6	43.4	81.2	83.8	62.3	35.2	66.1
Full-Hist-10	75.5	19.6	95.5	93.1	55.9	40.0	63.3
OppGabor	85.1	30.7	83.6	88.9	53.6	34.4	62.7
<i>CNN-based features</i>							
ResNet-101-FC	94.2	51.9	97.1	98.3	83.9	75.3	83.4
ResNet-152-FC	94.0	53.3	97.1	98.3	83.5	75.7	83.6
ResNet-50-FC	94.1	52.6	95.8	98.1	84.2	75.0	83.3
VGG-VD-19-FC	91.0	40.6	96.5	97.2	80.4	73.2	79.8
VGG-VD-16-FC	90.7	39.2	96.2	97.0	81.4	73.3	79.6

Datasets: 1) CURET-61-S-M; 2) Fabrics-1968-S-M; 3) KTH-TIPS-10-S-M; 4) KTH-TIPS2b-10-S-M; 5) LMT-94-S-M; 6) RDAD-27-S-M.

Another interesting outcome is the relative performance of the descriptors within the two classes of methods. The ranking of the CNN-based features was rather stable across all the experiments, with the three ResNet models invariably sitting in the first places of the standings. Conversely, hand-crafted descriptors showed a higher degree of variability: LBP colour variants (e.g., OCLBP, IOCLBP, and LCVBP) for instance—which were among the best methods on the whole—did not perform well under variable illumination (as one would expect) and were in fact surpassed by grey-scale methods (e.g., SIFT and ILBP).

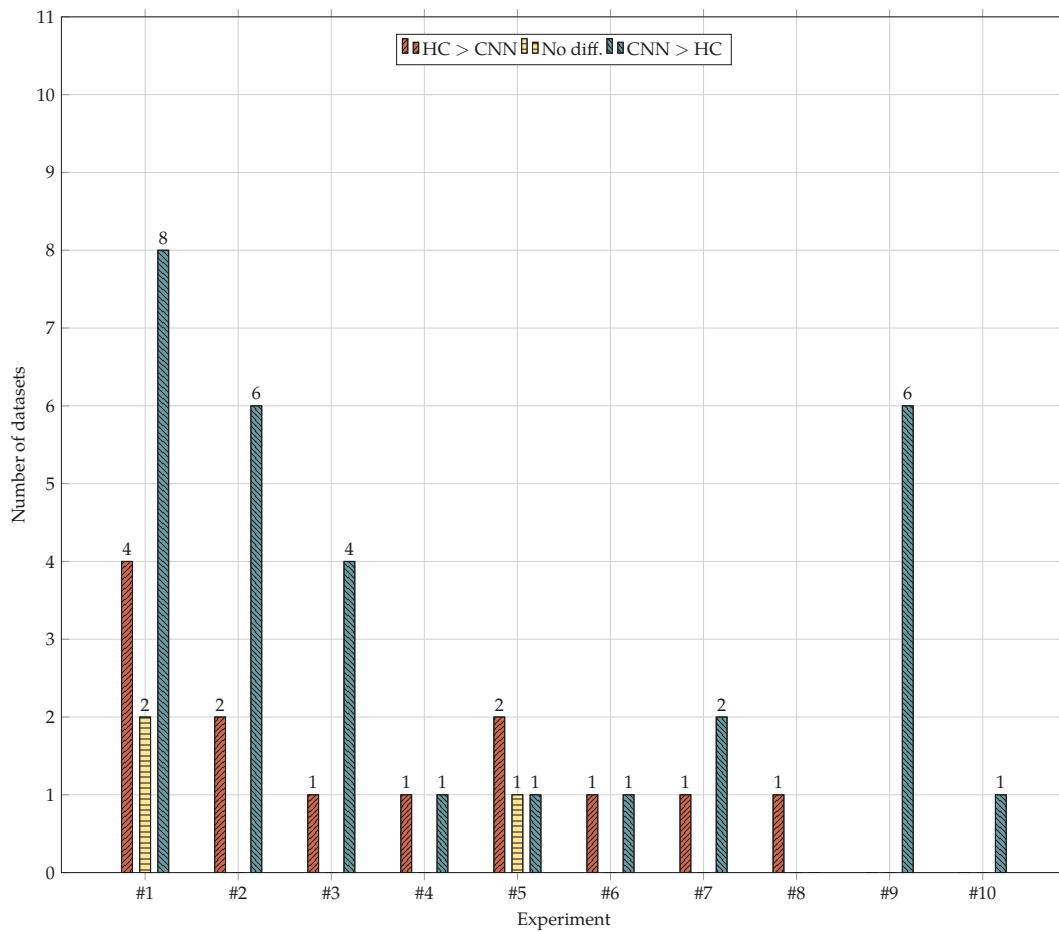


Figure 16. Number of datasets on which: hand-crafted methods performed significantly better than CNNs (“HC > CNN”); there was no significant difference (“No diff.”) and CNNs performed significantly better than hand-crafted methods (“CNN > HC”). see also Tables 6–13.

Table 16. Results of Experiment #10: non-stationary textures with multiple variations. Figures report overall accuracy by dataset. Boldface denotes the highest value; underline signals a statistically-significant difference between the best hand-crafted and the best CNN-based descriptor. Descriptors are listed in ascending order of by-experiment rank (best first).

Descriptor	Dataset
	1
<i>Hand-crafted image descriptors</i>	
SIFT-BoVW	31.6
VZ-MR8	27.8
SIFT-VLAD	27.1
IPBC-J	22.7
ILBP	21.0
<i>CNN-based features</i>	
ResNet-50-FC	60.8
ResNet-152-FC	60.4
ResNet-101-FC	60.3
VGG-VD-19-FC	56.2
VGG-VD-16-FC	56.2

Datasets: 1) DTD-47-NS-M.

6.2. Computational Demand

Table 17 reports, for each image descriptor, the average Feature Extraction time per image (FE) and the average Classification time per subdivision into the training and test set (CL). The figures were recorded from Experiment #1. For a fair comparison, all the features were computed using the CPU only (no GPU acceleration was used). To facilitate a comparative assessment, we subdivided the whole population into quartiles (columns Q_{FE} and Q_{CL} of the table). On the whole, the results indicate that the best performing hand-crafted descriptors (e.g., OCLBP, LCVBP and ICM) were generally slower than the CNN-based methods in the feature extraction step; in the classification step, however, the situation was inverted in favour of the hand-crafted descriptors due to the lower dimensionality of these methods.

Table 17. Computational demand: FE = average Feature Extraction time per image, CL = average Classification time per problem; Q_{FE} and Q_{CL} are the corresponding quartiles. Values are in seconds. Note: features’ extraction and classification from the DenseNet-161 and DenseNet-201 models were carried out on a machine different from the one used for all the other descriptors. Consequently, computing times for DenseNets are not directly comparable to those of the other descriptors.

<i>Hand-Crafted Image Descriptors</i>					<i>CNN-Based Features</i>				
Abbreviation	FE	Q_{FE}	CL	Q_{CL}	Abbreviation	FE	Q_{FE}	CL	Q_{CL}
<i>Purely spectral</i>					Caffe-Alex-FC	0.089	I	0.611	III
Mean	0.043	I	0.007	I	Caffe-Alex-BoVW	0.151	I	0.676	IV
Mean + Std	0.054	I	0.007	I	Caffe-Alex-VLAD	0.101	I	0.626	IV
Mean + Moms	0.152	I	0.009	I	Caffe-Alex-BoVW	0.151	I	0.676	IV
Quartiles	0.063	I	0.007	I	DenseNet-161-FC	1.060	*	2.565	*
Marg.-Hists-256	0.103	I	0.114	II	DenseNet-161-BoVW	1.462	*	2.647	*
Full-Hist-10	0.168	II	0.147	II	DenseNet-201-FC	0.858	*	0.924	*
<i>Grey-scale texture</i>					DenseNet-201-BoVW	0.986	*	0.973	*
CLBP	0.564	II	0.053	II	GoogLeNet-FC	0.717	III	0.146	II
GLBP	0.272	II	0.021	I	GoogLeNet-BoVW	0.709	III	0.150	II
ILBP	0.329	II	0.036	I	ResNet-50-FC	0.736	III	0.285	III
LBP	0.267	II	0.021	I	ResNet-50-BoVW	0.800	III	0.333	III
LTP	0.668	III	0.036	I	ResNet-101-FC	1.416	III	0.289	III
TS	0.606	III	0.406	III	ResNet-101-BoVW	1.461	IV	0.338	III
GLCM	0.789	III	0.014	I	ResNet-152-FC	2.068	IV	0.285	III
GLCM ^{DFT}	0.812	III	0.015	I	ResNet-152-BoVW	2.112	IV	0.332	III
Gabor	1.396	III	0.016	I	VGG-Face-FC	0.381	II	0.614	III
Gabor _{cn}	1.455	IV	0.016	I	VGG-Face-BoVW	0.459	II	0.699	IV
Gabor ^{DFT}	1.395	III	0.016	I	VGG-Face-VLAD	0.383	II	0.620	IV
Gabor _{cn} ^{DFT}	1.457	IV	0.016	I	VGG-F-FC	0.084	I	0.614	III
IPBC-J	4.300	IV	0.681	IV	VGG-F-BoVW	0.133	I	0.687	IV
HOG	0.142	I	0.115	II	VGG-F-VLAD	0.092	I	0.612	III
VZ-MR8	4.366	IV	0.679	IV	VGG-M-FC	0.160	I	0.614	III
SIFT-BoVW	10.064	IV	0.676	IV	VGG-M-BoVW	0.223	II	0.689	IV
SIFT-VLAD	0.617	III	0.612	III	VGG-M-VLAD	0.154	I	0.613	III
WSF + WCF ⁽¹⁾	0.953	III	0.018	I	VGG-S-FC	0.147	I	0.616	IV
WSF + WCF ⁽²⁾	1.018	III	0.018	I	VGG-S-BoVW	0.269	II	0.688	IV
<i>Colour texture</i>					VGG-S-VLAD	0.155	I	0.614	III
OCLBP	1.183	III	0.097	II	VGG-VD-16-FC	0.382	II	0.615	IV
IOCLBP	1.580	IV	0.188	II	VGG-VD-16-BoVW	0.463	II	0.679	IV
LCVBP	2.194	IV	0.067	II	VGG-VD-16-VLAD	0.388	II	0.612	III
LBP + LCC	0.918	III	0.130	II	VGG-VD-19-FC	0.425	II	0.612	III
ICM	4.495	IV	0.057	II	VGG-VD-19-BoVW	0.508	II	0.679	IV
ICM ^{DFT}	4.508	IV	0.056	II	VGG-VD-19-VLAD	0.437	II	0.611	III
OppGabor	5.427	IV	0.094	II					
OppGabor _{cn}	5.614	IV	0.094	II					
OppGabor ^{DFT}	5.440	IV	0.094	II					
OppGabor _{cn} ^{DFT}	5.617	IV	0.095	II					

⁽¹⁾ Haar wavelet; ⁽²⁾ Bi-orth.wavelet.

7. Conclusions

We have compared the effectiveness and computational workload of traditional, hand-crafted descriptors against off-the-shelf CNN-based features for colour texture classification under ideal and realistic conditions. On average, the experiments confirmed the superiority of deep networks, albeit with some interesting exceptions. Specifically, hand-crafted descriptors still proved better than CNN-based features when there was little intra-class variability or where this could be modelled explicitly (e.g., rotations). The reverse was true when there was significant intra-class variability—whether due to the intrinsic structure of the images and/or to changes in the imaging conditions—and in general in all the other cases.

Of the three aggregation techniques used for extracting features via pre-trained CNN (i.e., FC, BoVW and VLAD), the first outperformed the other two in all the conditions considered. This finding is in agreement with the results recently published by Cusano et al. [17], but differs from those presented by Cimpoi [16] in which VLAD (and to some extent BoVW) performed either better or at least as well as FC. Note, however, that in our comparison, we kept the number of features approximately equal for the three methods, whereas in [16], VLAD's feature vectors were significantly longer than FC and BoVW. Furthermore, consider that in our experiments, the aggregation was performed over an *external*—dataset-independent—dictionary, whereas [16] used dataset-specific (*internal*) dictionaries. Incidentally, it is worth noting that the FC configuration is the only one that allows a genuine off-the-shelf reuse of the networks in a strict sense, only requiring a resizing of the input images to fit the input field of the net.

Among the hand-crafted descriptors, colour LBP variants such as OCLBP, IOCLBP and LCVBP gave the best results under stable illumination, whereas dense SIFT proved the most effective method in the presence of illumination changes. Pure colour descriptors (i.e., full and marginal colour histograms) were the best methods to deal with variations in scale.

On the other hand, the performance of CNN-based features was rather stable across all the datasets and experiments, with the three ResNet models emerging as the best descriptors in nearly all experimental conditions.

As for the computational cost, the best CNN-based features were approximately as fast to compute as their hand-crafted counterparts (Table 17). The feature vectors, however, are at least twice as long (Tables 2 and 3), which implies higher computational demand in the classification step.

Finally, an interesting and rather curious result is the high affinity that emerged between local binary patterns and the Outex database: in most of the experiments in which this dataset was involved, the best descriptor was always an LBP variant, a finding that did not go unnoticed by other authors either [16].

Supplementary Materials: The datasets' file names, class labels and links to the original images are available at the Supplementary material. Feature extraction and classification were based on CATAcOMB (Colour And Texture Analysis toolbox for Matlab), which is provided in the CATAcOMB.zip file. An on-line version of the toolbox will also be made available at <https://bitbucket.org/biancovic/catacomb> upon publication.

Author Contributions: Conceptualization, R.-B.C., F.B., P.N. and F.S.; methodology, R.-B.C., F.B., P.N. and F.S.; software, R.-B.C., F.B. and P.N.; validation, R.-B.C., F.B., P.N., F.D.M. and F.S.; resources, F.B., F.D.M. and P.N.; data curation, R.-B.C. and F.B.; writing, original draft preparation R.-B.C. and F.B.; writing, review and editing, R.-B.C., F.B., F.D.M., P.N. and F.S.; funding acquisition, F.B. and F.D.M.

Funding: This research was partially supported by the Department of Engineering at the Università degli Studi di Perugia, Italy, through the Fundamental Research Grants Scheme 2017, by the Italian Ministry of Education, University and Research (MIUR) within the Annual Individual Funding Scheme for Fundamental Research (FFABR2017) and by the Spanish Government under Project AGL2014-56017-R.

Acknowledgments: The authors wish to thank Richard Bormann (Fraunhofer-Institut für Produktionstechnik und Automatisierung, Germany) for the RDAD dataset and Luiz Eduardo S. Oliveira (Federal University of Paraná, Brazil) for the ForestSpecies dataset.

Conflicts of Interest: The authors declare no conflict of interest. The funders mentioned in the 'Funding' section had no role whatsoever in the design of the study; the collection, analyses, or interpretation of the data; the writing of the manuscript; nor the decision to publish the results.

References

1. Liu, L.; Fieguth, P.; Guo, Y.; Wang, X.; Pietikäinen, M. Local binary features for texture classification: Taxonomy and experimental study. *Pattern Recognit.* **2017**, *62*, 135–160. [CrossRef]
2. Liu, L.; Chen, J.; Fieguth, P.; Zhao, G.; Chellappa, R.; Pietikäinen, M. From BoW to CNN: Two Decades of Texture Representation for Texture Classification. *Int. J. Comput. Vis.* **2019**, *127*, 74–109. [CrossRef]
3. Humeau-Heurtier, A. Texture feature extraction methods: A survey. *IEEE Access* **2019**. [CrossRef]
4. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Volume 2, pp. 1097–1105.
5. Razavian, A.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN features off-the-shelf: An astounding baseline for recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 512–519.
6. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 5th International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
7. Parkhi, O.; Vedaldi, A.; Zissermann, A. Deep face recognition. In Proceedings of the British Machine Vision Conference, Swansea, UK, 7–10 September 2015.
8. Nam, G.P.; Choi, H.; Cho, J.; Kim, I.J. PSI-CNN: A Pyramid-Based Scale-Invariant CNN Architecture for Face Recognition Robust to Various Image Resolutions. *Appl. Sci.* **2018**, *8*, 1561. [CrossRef]
9. Hertel, L.; Barth, E.; Kaster, T.; Martinetz, T. Deep convolutional neural networks as generic feature extractors. In Proceedings of the International Joint Conference on Neural Networks, Killarney, Ireland, 12–17 July 2015.
10. Mäenpää, T.; Pietikäinen, M. Classification with color and texture: Jointly or separately? *Pattern Recognit.* **2004**, *37*, 1629–1640. [CrossRef]
11. Bianconi, F.; Harvey, R.; Southam, P.; Fernández, A. Theoretical and experimental comparison of different approaches for color texture classification. *J. Electron. Imaging* **2011**, *20*, 043006:1–043006:17. [CrossRef]
12. Kandaswamy, U.; Schuckers, S.A.; Adjero, D. Comparison of Texture Analysis Schemes Under Nonideal Conditions. *IEEE Trans. Image Process.* **2011**, *20*, 2260–2275. [CrossRef] [PubMed]
13. Qazi, I.; Alata, O.; Burie, J.; Moussa, A.; Maloigne, C.F. Choice of a pertinent color space for color texture characterization using parametric spectral analysis. *Pattern Recognit.* **2011**, *44*, 16–31. [CrossRef]
14. Cernadas, E.; Fernández-Delgado, M.; González-Rufino, E.; Carrión, P. Influence of normalization and color space to color texture classification. *Pattern Recognit.* **2017**, *61*, 120–138. [CrossRef]
15. Cimpoi, M.; Maji, S.; Vedaldi, A. Deep filter banks for texture recognition and segmentation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3828–3836.
16. Cimpoi, M.; Maji, S.; Kokkinos, I.; Vedaldi, A. Deep Filter Banks for Texture Recognition, Description, and Segmentation. *Int. J. Comput. Vis.* **2016**, *118*, 65–94. [CrossRef] [PubMed]
17. Cusano, C.; Napoletano, P.; Schettini, R. Combining multiple features for color texture classification. *J. Electron. Imaging* **2016**, *25*, 061410. [CrossRef]
18. Napoletano, P. Hand-Crafted vs. Learned Descriptors for Color Texture Classification. In *Proceedings of the 6th Computational Color Imaging Workshop (CCIW'17)*; Bianco, S., Schettini, R., Tominaga, S., Treméau, A., Eds.; Lecture Notes in Computer Science; Springer: Milan, Italy, 2017; Volume 10213, pp. 259–271.
19. Petrou, M.; García Sevilla, P. *Image Processing. Dealing with Texture*; Wiley Interscience: Chichester, UK, 2006.
20. Burghouts, G.J.; Geusebroek, J.M. Material-specific adaptation of color invariant features. *Pattern Recognit. Lett.* **2009**, *30*, 306–313. [CrossRef]
21. Amsterdam Library of Textures. Available online: http://aloi.science.uva.nl/public_alot/ (accessed on 11 January 2017).
22. Brodatz, P. *Textures: A Photographic Album for Artists and Designers*; Dover: New York, NY, USA, 1966.
23. Colored Brodatz Texture Database. Available online: http://multibandtexture.recherche.usherbrooke.ca/colored_brodatz.html (accessed on 11 January 2017).
24. Dana, K.J.; van Ginneken, B.; Nayar, S.K.; Koenderink, J.J. Reflectance and Texture of Real-World Surfaces. *ACM Trans. Graph.* **1999**, *18*, 1–34. [CrossRef]

25. CURET: Columbia-Utrecht Reflectance and Texture Database. Available online: <http://www.cs.columbia.edu/CAVE/software/curet/index.php> (accessed on 25 January 2017).
26. Visual Geometry Group. CURET: Columbia-Utrecht Reflectance and Texture Database. Available online: <http://www.robots.ox.ac.uk/~vgg/research/textclass/setup.html> (accessed on 26 January 2017).
27. Oxholm, G.; Bariya, P.; Nishino, K. The scale of geometric texture. In *Proceedings of the 12th European Conference on Computer Vision (ECCV 2012)*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Florence, Italy, 2012; Volume 7572, pp. 58–71.
28. Drexel Texture Database. Available online: <https://www.cs.drexel.edu/~kon/texture/> (accessed on 11 January 2017).
29. Cimpoi, M.; Maji, S.; Kokkinos, I.; Mohamed, S.; Vedaldi, A. Describing textures in the wild. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 23–28 June 2014; pp. 3606–3613.
30. Describable Texture Database (DTD). Available online: <https://www.robots.ox.ac.uk/~vgg/data/dtd/> (accessed on 19 January 2019).
31. Kampouris, C.; Zafeiriou, S.; Ghosh, A.; Malassiotis, S. Fine-grained material classification using micro-geometry and reflectance. In *Proceedings of the 14th European Conference on Computer Vision (ECCV 2016)*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Amsterdam, The Netherlands, 2016; Volume 9909, pp. 778–792.
32. The Fabrics Dataset. Available online: <http://ibug.doc.ic.ac.uk/resources/fabrics/> (accessed on 11 January 2017).
33. Martins, J.; Oliveira, L.S.; Nigkoski, S.; Sabourin, R. A database for automatic classification of forest species. *Mach. Vis. Appl.* **2013**, *24*, 567–578. [[CrossRef](#)]
34. ForestSpecies Database. Available online: <http://web.inf.ufpr.br/vri/image-and-videos-databases/forest-species-database> (accessed on 11 January 2017).
35. Hayman, E.; Caputo, B.; Fritz, M.; Eklundh, J. On the significance of real-world conditions for material classification. In *Proceedings of the 8th European Conference on Computer Vision (ECCV 2004)*, Prague, Czech Republic, 11–14 May 2002; Volume 3024, pp. 253–266.
36. The KTH-TIPS and KTH-TIPS2 Image Databases. Available online: <http://www.nada.kth.se/cvap/databases/kth-tips/download.html> (accessed on 11 January 2017).
37. Caputo, B.; Hayman, E.; Mallikarjuna, P. Class-specific material categorisation. In *Proceedings of the 10th International Conference on Computer Vision (ICCV)*, Beijing, China, 17–20 October 2005; Volume 2, pp. 1597–1604.
38. Kylberg, G. Automatic Virus Identification Using TEM. Image Segmentation and Texture Analysis. Ph.D. Thesis, Faculty of Science and Technology, University of Uppsala, Uppsala, Sweden, 2014.
39. Kylberg Sintorn Rotation Dataset. Available online: <http://www.cb.uu.se/~gustaf/KylbergSintornRotation/> (accessed on 11 January 2017).
40. Strese, M.; Schuwerk, C.; Iepure, A.; Steinbach, E. Multimodal Feature-Based Surface Material Classification. *IEEE Trans. Haptics* **2017**, *10*, 226–239. [[CrossRef](#)] [[PubMed](#)]
41. LMT Haptic Texture Database. Available online: <http://www.lmt.ei.tum.de/downloads/texture/> (accessed on 25 January 2017).
42. Bianconi, F.; Bello-Cerezo, R.; Fernández, A.; González, E. On comparing colour spaces from a performance perspective: application to automated classification of polished natural stones. In *New Trends in Image Analysis and Processing—ICIAP 2015 Workshops*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9281, pp. 71–78.
43. MondialMarmi: A Collection of Images of Polished Natural Stones for Colour and Texture Analysis, v2.0. Available online: http://dismac.dii.unipg.it/mm/ver_2_0/index.html (accessed on 19 January 2019).
44. Abdelmounaime, S.; Dong-Chen, H. New Brodatz-based Image Databases for Grayscale Color and Multiband Texture Analysis. *ISRN Mach. Vis.* **2013**, *2013*, 876386. [[CrossRef](#)]
45. Multiband Texture Database. Available online: http://multibandtexture.recherche.usherbrooke.ca/multi_band_more.html (accessed on 12 January 2017).

46. Normalized Brodatz's Texture Database. Available online: http://multibandtexture.recherche.usherbrooke.ca/normalized_brodatz.html (accessed on 23 February 2017).
47. Porebski, A.; Vandenbroucke, N.; Macaire, L.; Hamad, D. A new benchmark image test suite for evaluating color texture classification schemes. *Multimed. Tools Appl. J.* **2014**, *70*, 543–556. [[CrossRef](#)]
48. New [BarkTex] Benchmark Image Test Suite for Evaluating Color Texture Classification Schemes. Available online: https://www-lisic.univ-littoral.fr/~porebski/BarkTex_image_test_suite.html (accessed on 12 January 2017).
49. Palm, C.; Lehmann, T.M. Classification of color textures by Gabor filtering. *Mach. Graph. Vis.* **2002**, *11*, 195–219.
50. Ojala, T.; Pietikäinen, M.; Mäenpää, T.; Viertola, J.; Kyllönen, J.; Huovinen, S. Outex—New Framework for Empirical Evaluation of Texture Analysis Algorithms. In *Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02)*; IEEE Computer Society: Quebec City, QC, Canada, 2002; Volume 1, pp. 701–706.
51. Outex Texture Database. Available online: <http://www.outex oulu.fi/> (accessed on 12 January 2018).
52. Bianconi, F.; Fernández, A.; González, E.; Saetta, S. Performance analysis of colour descriptors for parquet sorting. *Expert Syst. Appl.* **2013**, *40*, 1636–1644. [[CrossRef](#)]
53. Parquet Dataset. Available online: <http://dismac.dii.unipg.it/parquet/index.html> (accessed on 12 January 2017).
54. Casanova, D.; Sá, J.J.; Bruno, O. Plant leaf identification using Gabor wavelets. *Int. J. Imaging Syst. Technol.* **2009**, *19*, 236–246. [[CrossRef](#)]
55. 1200Tex Dataset. Available online: <https://scg.ifsc.usp.br/dataset/1200Tex.php/> (accessed on 11 January 2019).
56. Bormann, R.; Esslinger, D.; Hundsdörfer, D.; Högele, M.; Vincze, M. Texture characterization with semantic attributes: Database and algorithm. In *Proceedings of the 47th International Symposium on Robotics (ISR 2016)*, Munich, Germany, 21–22 June 2016; pp. 149–156.
57. Cusano, C.; Napoletano, P.; Schettini, R. Evaluating color texture descriptors under large variations of controlled lighting conditions. *J. Opt. Soc. Am. A* **2016**, *33*, 17–30. [[CrossRef](#)] [[PubMed](#)]
58. RawFoot DB: Raw Food Texture Database. Available online: <http://projects.ivl.disco.unimib.it/minisites/rawfoot/> (accessed on 12 January 2017).
59. Salzburg texture Image Database (STex). Available online: <http://www.wavelab.at/sources/STex/> (accessed on 12 January 2017).
60. Backes, A.; Casanova, D.; Bruno, O. Color texture analysis based on fractal descriptors. *Pattern Recognit.* **2012**, *45*, 1984–1992. [[CrossRef](#)]
61. USPTex Database. Available online: <http://fractal.ifsc.usp.br/dataset/USPtex.php> (accessed on 19 January 2019).
62. VisTex Reference Textures. Available online: <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/Images/Reference/> (accessed on 12 January 2017).
63. López, F.; Valiente, J.M.; Prats, J.M.; Ferrer, A. Performance evaluation of soft color texture descriptors for surface grading using experimental design and logistic regression. *Pattern Recognit.* **2008**, *41*, 1744–1755. [[CrossRef](#)]
64. Bello-Cerezo, R.; Bianconi, F.; Fernández, A.; González, E.; Di Maria, F. Experimental comparison of color spaces for material classification. *J. Electron. Imaging* **2016**, *25*, 061406. [[CrossRef](#)]
65. Pietikäinen, M.; Nieminen, S.; Marszalec, E.; Ojala, T. Accurate Color Discrimination with Classification Based on Features Distributions. In *Proceedings of the 13th International Conference on Pattern Recognition (ICPR'96)*, Vienna, Austria, 25–29 August 1996; Volume 3, pp. 833–838.
66. Swain, M.J.; Ballard, D.H. Color Indexing. *Int. J. Comput. Vis.* **1991**, *7*, 11–32. [[CrossRef](#)]
67. Fernández, A.; Álvarez, M.X.; Bianconi, F. Texture description through histograms of equivalent patterns. *J. Math. Imaging Vis.* **2013**, *45*, 76–102. [[CrossRef](#)]
68. Guo, Z.; Zhang, L.; Zhang, D. A Completed Modeling of Local Binary Pattern Operator for Texture Classification. *IEEE Trans. Image Process.* **2010**, *19*, 1657–1663.
69. He, Y.; Sang, N. Robust Illumination Invariant Texture Classification Using Gradient Local Binary Patterns. In *Proceedings of the 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping*, Xiamen, China, 10–12 January 2011; pp. 1–6.
70. Jin, H.; Liu, Q.; Lu, H.; Tong, X. Face detection using improved LBP under Bayesian framework. In *Proceedings of the 3rd International Conference on Image and Graphics*, Hong Kong, China, 18–20 December 2004; pp. 306–309.

71. Ojala, T.; Pietikäinen, M.; Mäenpää, T. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [[CrossRef](#)]
72. Tan, X.; Triggs, B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.* **2010**, *19*, 1635–1650.
73. He, D.C.; Wang, L. Texture Unit, Texture Spectrum, And Texture Analysis. *IEEE Trans. Geosci. Remote Sens.* **1990**, *28*, 509–512.
74. González, E.; Bianconi, F.; Fernández, A. An investigation on the use of local multi-resolution patterns for image classification. *Inf. Sci.* **2016**, *361*–362, 1–13.
75. Pardo-Balado, J.; Fernández, A.; Bianconi, F. Texture classification using rotation invariant LBP based on digital polygons. In *New Trends in Image Analysis and Processing—ICIAP 2015 Workshops*; Lecture Notes in Computer Science; Murino, V., Puppo, E., Eds.; Springer: Basel, Switzerland, 2015; Volume 9281, pp. 87–94.
76. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *3*, 610–621. [[CrossRef](#)]
77. Bianconi, F.; Fernández, A. Rotation invariant co-occurrence features based on digital circles and discrete Fourier transform. *Pattern Recognit. Lett.* **2014**, *48*, 34–41. [[CrossRef](#)]
78. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), San Diego, CA, USA, 20–26 June 2005; Volume I, pp. 886–893.
79. Varma, M.; Zisserman, A. A statistical approach to material classification using image patch exemplars. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2032–2047. [[CrossRef](#)] [[PubMed](#)]
80. Lahajnar, F.; Kovacic, S. Rotation-invariant texture classification. *Pattern Recognit. Lett.* **2003**, *24*, 1151–1161. [[CrossRef](#)]
81. Arivazhagan, S.; Ganesan, L. Texture classification using wavelet transform. *Pattern Recognit. Lett.* **2003**, *24*, 1513–1521. [[CrossRef](#)]
82. Varma, M.; Zisserman, A. A Statistical Approach to Texture Classification from Single Images. *Int. J. Comput. Vis.* **2005**, *62*, 61–81. [[CrossRef](#)]
83. Arvis, V.; Debain, C.; Berducat, M.; Benassi, A. Generalization of the cooccurrence matrix for colour images: Application to colour texture classification. *Image Anal. Stereol.* **2004**, *23*, 63–72. [[CrossRef](#)]
84. Palm, C. Color texture classification by integrative Co-occurrence matrices. *Pattern Recognit.* **2004**, *37*, 965–976. [[CrossRef](#)]
85. Cusano, C.; Napoletano, P.; Schettini, R. Combining local binary patterns and local color contrast for texture classification under varying illumination. *J. Opt. Soc. Am. A: Opt. Image Sci. Vis.* **2014**, *31*, 1453–1461. [[CrossRef](#)] [[PubMed](#)]
86. Lee, S.; Choi, J.; Ro, Y.; Plataniotis, K. Local color vector binary patterns from multichannel face images for face recognition. *IEEE Trans. Image Process.* **2012**, *21*, 2347–2353. [[CrossRef](#)] [[PubMed](#)]
87. Mäenpää, T.; Pietikäinen, M. Texture Analysis with Local Binary Patterns. In *Handbook of Pattern Recognition and Computer Vision*, 3rd ed.; Chen, C.H., Wang, P.S.P., Eds.; World Scientific Publishing: Singapore, 2005; pp. 197–216.
88. Bianconi, F.; Bello-Cerezo, R.; Napoletano, P. Improved opponent color local binary patterns: An effective local image descriptor for color texture classification. *J. Electron. Imaging* **2018**, *27*, 011002. [[CrossRef](#)]
89. Jain, A.; Healey, G. A Multiscale Representation Including Opponent Color Features for Texture Recognition. *IEEE Trans. Image Process.* **1998**, *7*, 124–128. [[CrossRef](#)] [[PubMed](#)]
90. Huang, G.; Liu, Z.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
91. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
92. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

93. Chatfield, K.; Simonyan, K.; Vedaldi, A.; Zisserman, A. Return of the devil in the details: Delving deep into convolutional nets. In Proceedings of the British Machine Vision Conference 2014, Nottingham, UK, 1–5 September 2014.
94. Bianconi, F.; Fernández, A. A unifying framework for LBP and related methods. In *Local Binary Patterns: New Variants and Applications*; Brahmam, S., Jain, L.C., Nanni, L., Lumini, A., Eds.; Studies in Computational Intelligence; Springer: Berlin, Germany 2014; Volume 506, pp. 17–46.
95. Jégou, H.; Perronnin, F.; Douze, M.; Sánchez, J.; Pérez, P.; Schmid, C. Aggregating local image descriptors into compact codes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1704–1716. [[CrossRef](#)] [[PubMed](#)]
96. Texture King. Free Stock Textures, TextureKing. Available online: <http://www.textureking.com> (accessed on 14 December 2016).
97. Kanji, G. *100 Statistical Tests*, 3rd ed.; SAGE Publications Ltd.: London, UK, 2006.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).