Accepted version of paper published in International Conference on Multimedia Retrieval (ICMR), 2021.
DOI: https://doi.org/10.1145/3460426.3463645

A Denoising Convolutional Neural Network for Self-Supervised Rank Effectiveness Estimation on Image Retrieval

Lucas Pascotti Valem and Daniel Carlos Guimarães Pedronette
Department of Statistics, Applied Math. and Computing, São Paulo State University (UNESP)
Rio Claro, SP, Brazil
lucas.valem@unesp.br,daniel.pedronette@unesp.br

ABSTRACT

Image and multimedia retrieval has established as a prominent task in an increasingly digital and visual world. Mainly supported by decades of development on hand-crafted features and the success of deep learning techniques, various different feature extraction and retrieval approaches are currently available. However, the frequent requirements for large training sets still remain as a fundamental bottleneck, especially in real-world and large-scale scenarios. In the scarcity or absence of labeled data, choosing what retrieval approach to use became a central challenge. A promising strategy consists in to estimate the effectiveness of ranked lists without requiring any groundtruth data. Most of the existing measures exploit statistical analysis of the ranked lists and measure the reciprocity among lists of images in the top positions. This work innovates by proposing a new and self-supervised method for this task, the Deep Rank Noise Estimator (DRNE). An algorithm is presented for generating synthetic ranked list data, which is modeled as images and provided for training a Convolutional Neural Network that we propose for effectiveness estimation. The proposed model is a variant of the DnCNN (Denoiser CNN), which intends to interpret the incorrectness of a ranked list as noise, which is learned by the network. Our approach was evaluated on 5 public image datasets and different tasks, including general image retrieval and person re-ID. We also exploited and evaluated the complementary between the proposed approach and related rank-based approaches through fusion strategies. The experimental results showed that the proposed method is capable of achieving up to 0.88 of Pearson correlation with MAP measure in general retrieval scenarios and 0.74 in person re-ID scenarios.

CCS CONCEPTS

• **Information systems** → **Information retrieval**.

KEYWORDS

content-based image retrieval, effectiveness estimation, query performance prediction, unsupervised learning, self-supervised learning, convolutional neural networks, denoising

ICMR '21, August 21–24, 2021, Taipei, Taiwan

ACM ISBN 978-1-4503-8463-6/21/08…$15.00
https://doi.org/10.1145/3460426.3463645

ACM Reference Format:
Lucas Pascotti Valem and Daniel Carlos Guimarães Pedronette. 2021. A Denoising Convolutional Neural Network for Self-Supervised Rank Effectiveness Estimation on Image Retrieval. In *Proceedings of the 2021 International Conference on Multimedia Retrieval (ICMR '21), August 21–24, 2021, Taipei, Taiwan.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3460426.3463645

1 INTRODUCTION

The ubiquitous access to image acquisition devices and widespread facilities on storage and sharing triggered a consistent increase of image collections and related applications [70]. The task of retrieving images according to their visual content addressed by CBIR (Content-based Image Retrieval) systems [54] is a well studied problem which still attracts a lot of attention of the scientific community [70]. Overall, for a query image, the most similar dataset images are ranked according to their similarity, in decreasing order. Thus, the retrieval effectiveness depends upon the similarity measurement between images, which ideally should be discriminative and robust [13]. It is inherently a challenging task, once there can be many interpretations for a same image and it may depend on the context of the search.

Originally, the evolution of CBIR systems was mainly supported by the development of image descriptors [54], defined by a feature extraction algorithm and a distance function. In order to make the retrieval robust to geometric and photometric variations, the image content in terms of the visual properties (color, texture, shape) is represented through a feature vector [13]. The idea is that a feature vector extracted for one image can be treated as its "fingerprint" and can ease up the search process. More formally, each image can be represented as a point in a high-dimensional space. The similarity measurement can be performed based on the distance between the points, computed by a distance function (often defined by the Euclidean distance).

For approximately a decade, deep learning have emerged and gave rise to a change of direction in feature representation research, from hand-engineering to learning-based [45]. The hierarchical feature representation design given by deep learning models are effective on learning the abstract features from data which are important for that dataset and application. In practice, Convolutional Neural Networks (CNN) features have achieved high-effective results on retrieval tasks [13]. However, deep learning approaches often require large amount of labeled data for training or rely on transfer learning strategies. In this scenario, the availability of label data for training deep learning models represents an important bottleneck, especially in real-world and large-scale scenarios.

Ranging between hand-crafted and deep learning approaches, a myriad of different feature extraction and retrieval approaches have

been proposed [67, 70]. Therefore, automatically choosing the most effective to use or fuse in each retrieval scenario became a central challenge. Many approaches were proposed based on labeled data and learning models to infer the effectiveness of each visual feature in order to select and fuse them [45]. Even in supervised scenarios, to perform an effective selection represents a difficult task. Thus, how to effectively select the visual features in an unsupervised way, is even challenger, once there is no information about effectiveness of individual features [55].

The possibility of estimating the effectiveness of a set of ranked lists computed by a given visual feature without the use of any labeled data is a very challenging but also very promising approach [42, 46]. Most of query performance prediction methods are based on supervised approaches [39]. Some few approaches addressed the problem in an unsupervised scenario, as the Authority Score [43] and Reciprocal Score [40]. Despite of the significant results, such measures are mainly based on graph-based formulations of ranking information and do not exploit deep learning models.

In this work, we propose a new method to estimate the effectiveness of ranked lists in a self-supervised fashion, the Deep Rank Noise Estimator (DRNE). The idea is to predict the quality of results generated by CBIR systems without labeled data. We propose a new model architecture based on a well known denoiser, which has results comparable to the state-of-the-art, the DnCNN [64] (Denoiser CNN). To keep the entire workflow unsupervised, we trained the model with synthetic data. In order to create such data, we emulate the behavior of real visual features with different degrees of effectiveness. Based on the generated data, the ranked lists are converted to images according to a strategy inspired by [44] and used to train the network, which interprets the incorrectness of a ranked list as noise. When the same representation is generated for real visual features, the network is able to estimate the noise and therefore the rank effectiveness.

To the best of our knowledge, this work is the first method which deals with the challenging task of unsupervised effectiveness estimation by using a denoising deep learning model. In this way, many research challenges in the context of this work can be highlighted, among them: *(i)* How can we accurately represent the ranked lists as images? *(ii)* What is the most efficient and scalable image representation? *(iii)* How to characterize the noise of these images as the incorrectness of the ranked lists? *(iv)* How to properly generate synthetic data in order to train the denoisers in a completely unsupervised way?

A broad experimental evaluation was conducted, considering 5 public image datasets, various visual features and different tasks, including general image retrieval and person re-ID tasks. The results are evaluated in terms of the correlation between the predicted effectiveness and the MAP score obtained by the visual feature. In addition, the complementary between the proposed approach and related rank-based approaches are exploited through fusion strategies. The results demonstrate that deep denoisers can be applied in order to predict the effectiveness of ranked lists computed by different features in distinct retrieval scenarios.

The remaining of the paper is organized as follows: Section 2 presents the related work, discussing the problem formulation and other effectiveness estimation measures; Section 3 describes the approach proposed in this paper; Section 4 shows the experimental protocol and the obtained results; finally, Section 5 states the conclusions and possible future works.

2 RELATED WORK

In this section, an overview with the most relevant related works is presented (Section 2.1) along with the rank model used in this paper (Section 2.2) and the effectiveness measures (Section 2.3) used as baselines.

2.1 Overview

Query performance prediction (QPP) [39] is a challenging task which consists in predicting, mostly post-query, the quality of results generated by an IR system. Initially proposed in text retrieval scenarios [9], such approaches also attracted the attention of the image retrieval research community [26, 39, 42, 46, 61], assuming a diverse taxonomy as query difficulty prediction [61], query difficulty estimation [26], and effectiveness estimation [42, 46].

This work innovates in comparison with related approaches [26, 39, 42, 46, 61] on modeling noise information to estimate effectiveness of retrieval results. However, one of the challenges of this work is how to properly represent the distance matrices or ranked lists, which are numeric data into a visual image that can be processed by denoising convolutional neural networks. In the literature, there are some works which proposed strategies to transform non-image data into images. In [48], the authors proposed different approaches for creating images from features vectors, like creating images of bar graphs and gray images where blacker pixels represent low distances and whiter pixels represent higher distance values.

There is also the DeepInsight approach [49] which is a very recent and promising technique. It consists into mapping all the features into a 2D space using a dimensionality reduction technique (e.g. t-SNE [33], kPCA). After the distribution is learned, the image is cropped according to its convex hull (the smallest rectangle where all the data points fit). The data points are represented according to the learned distribution and the differences in color are given according to differences in feature values. This approach can be used for different classification tasks where the datasets are not composed by images (e.g. text, audio, signals).

Regarding signal processing, which consists in a unidimensional data stream, a possible representation for analysing this data is the use of recurrence matrices [58]. This can be used to create images in order to analyse recurrent patterns between systems and functions. This technique provides a wide range of applications.

The proposed approach relies on the idea of noise removal from images which represent similarity information encoded in ranked lists, analogous to the approach performed in [44]. However, in this work, we train a denoising deep learning network, pairing the ranked list image to its MAP (Mean Average Precision) in order to obtain a score related to its effectiveness. In this way, we exploit the denoising network in a query performance prediction task.

Among the most relevant state-of-the-art deep denoisers, we can cite the DnCNN [64] (Denosing Convolutional Neural Network) which can learn noise patterns from pairs of clean and noisy images. The deep denoisers generally have the advantage of being capable of learning different noise patterns without requiring high execution times for parameter adjusting or image processing, like in most of the statistical approaches (e.g. BM3D [11]). There are also more recent approaches, like RDNN [65] (Residual Dense Neural Network) which was originally proposed for image super-resolution, but can also be employed for denoising tasks. More recently, there is the DRUnet [63], a variant of the UNet network employed for denoising.

The cited residual networks, besides more effective, generally tend to be less efficient regards time and more memory consuming when compared to DnCNN.

For training denoisers, the lack of clean image data to be used as groundtruth may be a challenge for certain applications like medical imaging and remote sensing [22, 35, 47, 50]. In this scenario, different training strategies were proposed. The Noise2Noise [24] and Noisier2Noise [35] approaches consist in the idea of training with pairs of noisy images, where the clean image can be predicted by learning common patterns in both images which are supposed to be present in the clean image. There is also the Noise2Void [21] where the learning process is done with only corrupted or noisy images, and the noisy pattern is learned considering the given dataset. There are also other strategies like the one based on Stein's unbiased risk estimator (SURE) [50] which proposes a MSE (Mean Squared Error) unsupervised estimation which can be used during training. Among the self-supervised strategies, there are some that implement a CNN with a "blind spot" in the receptive field of the network [22] and others that generate the groundtruth data using the most promising statistical methods (e.g. BM3D [11]) in order to train a network like DnCNN for example [50].

2.2 Problem Formulation

Let C=$\{x_1, x_2, \ldots, x_N\}$ be an image collection, where N denotes the collection size. Let us consider a retrieval task where, given a query image, returns a list of images from the collection C.

Formally, given a query image x_q, a ranker denoted by R_j computes a ranked list τ_q=$(x_1, x_2, \ldots, x_k)$ in response to the query. The ranked list τ_q can be defined as a permutation of the k-neighborhood set $\mathcal{N}(q)$, which contains the k most similar images to image x_q in the collection C. The permutation τ_q is a bijection from the set $\mathcal{N}(q)$ onto the set $[k] = \{1, 2, \ldots, k\}$. The $\tau_q(i)$ notation denotes the position (or rank) of image x_i in the ranked list τ_q.

The ranker R can be defined based on diverse approaches, including feature extraction or learning methods. In this paper, a feature-based approach is considered, defining R as a tuple (ϵ, ρ), where $\epsilon : C \rightarrow \mathbb{R}^d$ is a function that extracts a feature vector v_x from an image $x \in C$; and d: $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a distance function that computes the distance between two images according to their corresponding feature vectors. Formally, the distance between two images x_i, x_j is defined by $d(\epsilon(x_i), \epsilon(x_j))$. The notation $d(x_i, x_j)$ is used for readability purposes.

A ranked list can be computed by sorting images in a crescent order of distance. In terms of ranking positions we can say that, if image x_i is ranked before image x_j in the ranked list of image x_q, that is, $\tau_q(i) < \tau_q(j)$, then $d(q, i) \leq d(q, j)$. Taking every image in the collection as a query image x_q, a set of ranked lists $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_n\}$ can be obtained.

2.3 Unsupervised Effectiveness Estimation Measures

Among the various related works on effectiveness estimation for image retrieval [26, 39, 42, 46, 61], some are based on supervised learning [39] or specific representation models [26]. This section formally defines the measures most similar to our approach, defined in an unsupervised fashion and independent of retrieval model.

2.3.1 Authority Score. The Authority Measure [43] is used as effectiveness estimation measure by exploiting a graph representing the references among images defined in terms of ranked lists. Each image in top-k positions of the ranked list τ_q defines a node. For each image x_j in the top-k of τ_q, the ranked list τ_j is also analyzed. If there are images in common in ranked lists τ_q and τ_j, an edge is created. The Authority Score is computed based on the number of created edges. Therefore, the measure is based on the density of the graph and can be formally defined as follows:

$$\epsilon(\tau_q, k) = \frac{\sum_{u \in \mathcal{N}(q,k)} \sum_{v \in \mathcal{N}(u,k)} f_{in}(v, q)}{k^2}, \tag{1}$$

where $f_{in}(v, q)$ returns 1 if $\tau_q(v) \leq k$ and 0 otherwise. The value of ϵ is defined in the interval $[0, 1]$, achieving the greater score for a full connected graph at top-k positions.

2.3.2 Reciprocal Score. The Reciprocal Density Measure [40] is similar to the Authority, but it considers weights for each reciprocal neighbor:

$$R_s(\tau_q, k) = \frac{\sum_{i \in \mathcal{N}(q,k)} \sum_{j \in \mathcal{N}(i,k)} f_{in}(j, q) \times w_r(q, i) \times w_r(i, j)}{k^4}. \tag{2}$$

The function $f_{in}(j, q) \rightarrow \{0, 1\}$ returns 1 if $img_j \in \mathcal{N}(q, k)$. A weight is defined according to the function $w_r(q, i) = k + 1 - \tau_q(i)$. The higher the weight, more frequent is the occurrence of reciprocal neighbors in the first positions of the ranked list.

3 DEEP RANK NOISE ESTIMATOR (DRNE)

Our proposed strategy aims at computing effectiveness estimation measures for ranked lists without requiring any labeled data, in a self-supervised fashion. We name our method as Deep Rank Noise Estimator (DRNE). This approach can be summarized into three main steps:

(1) **Computing Synthetic Data**: they are used in order to simulate real scenarios for training the CNN, but without using any real label or groundtruth;
(2) **Ranked Lists as Images**: a strategy to represent ranked lists as images, since they are numerical data, they need to be converted to images to be provided as input to the CNN. Our approach for this is step is based on [44];
(3) **Effectiveness Estimation CNN**: the DnCNN [64] architecture was modified in order to be used as a effectiveness estimator for ranked lists, based on their "noise" level present on the images.

Each of the topics are better detailed and discussed in the following subsections.

3.1 Computing Synthetic Data

In general, training neural networks requires labeled data, which is not always easily available. With the objective of proposing a completely unsupervised training, we propose an algorithm to generate synthetic ranked lists that simulate real scenarios.

Our synthetic scenarios rely on the generation of a confusion matrix of probabilities. It is a squared CxC matrix where C is the number of virtual classes, present in the synthetic scenario. Being k the size of each virtual class, $C = N/k$, where N is the dataset size. The matrix is required to be symmetrical with all the values in the range $[0, 1]$, all the lines and columns are also required to sum to 1, to keep the consistency with the idea of probabilities. The

position (i, j) in this matrix corresponds to the probability of the elements of class i being mistaken by elements of class j. Following this reasoning, an element in the diagonal (position (i, j), where $i = j$) corresponds to the probability of an element being correctly attributed to its class. From this perspective, imposing restrictions for the values in the diagonal can increase or decrease the effectiveness of the ranked lists being generated. Figure 1 illustrates the similarities among classes, where the diagonal elements are highlighted in blue.

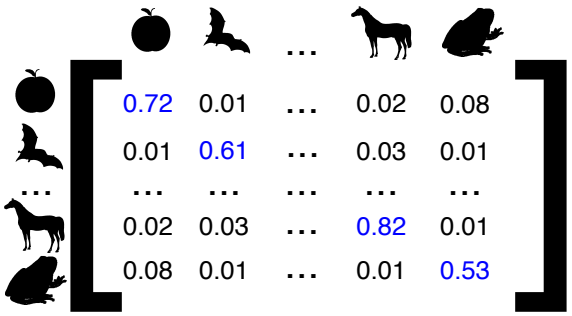


Figure 1: Illustration of a confusion matrix of probabilities between classes.

All ranked lists of the dataset share the same matrix for the generation of its elements. The elements are randomly generated according to the probabilities presented in the matrix. Since incorrect elements tend to be more random than the correct ones, we also generate a symmetrical similarity matrix to attribute weights for randomly select the elements that belong to the same class.

3.2 Ranked Lists as Images

The ranked lists consist in numerical data, where each value corresponds to the index of the image being ranked. In this work, we propose a model for transforming a ranked list into image data, based on what was proposed in [44].

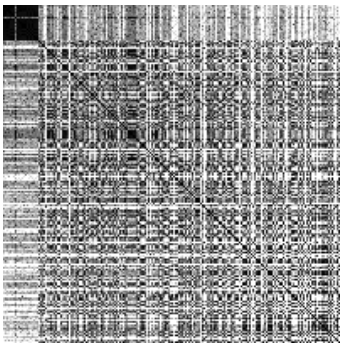
Given a pair of ranked lists τ_i and τ_j, a grayscale image can be modelled such that the pixel (p_x, p_y) is defined as the mean of the positions that the elements occur in both lists:

$$pixel(p_x, p_y) = (\tau_x(y) + \tau_y(x))/2, \quad (3)$$

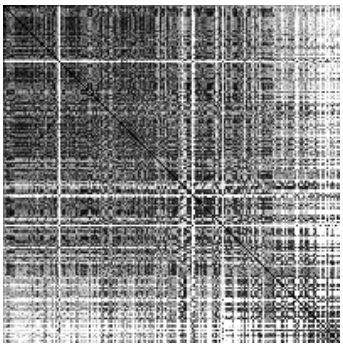
where $p_x = \tau_i(x)$ and $p_y = \tau_j(y)$.

For the purpose of this work, we always consider images of the same ranked list to the same ranked list (such that $\tau_i = \tau_j$), which produce symmetrical images. The positions with higher similarity are represented by darker pixels and the ones with lower similarity by brighter ones. Regarding the size of the images, we considered 200x200 in all the cases. The use of the same size of image for all the datasets shows the scalability potential of our method.

Figure 2 presents examples of images generated for synthetic ranked lists with different effectiveness levels. It can be seen that as the MAP (Mean Average Precision) increases, more blacker pixels tend to appear in the upper left corner of the image, for example. However, there are still many other features that can be analysed in this type of image, since there are multiple images for ranked lists with the same MAP.



(a) MAP = 0.953625



(b) MAP = 0.198827

Figure 2: Examples of synthetic data generated for different degrees of effectiveness ($k = 20$).

3.3 Effectiveness Estimation CNN

This work proposes a Convolutional Neural Network for estimating the effectiveness of ranked lists based on their image representations. The idea is that each ranked list image contains a certain level of noise, which is related to its effectiveness. Following this reasoning, the more effective a ranked list is, less noise is associated to it and vice versa.

Our method consists into apply the model to extract the noise and attribute a score which we expect that is related to the ranked list effectiveness. Figure 3 presents the model proposed and considered for all the experiments in this work. We modified the DnCNN [64] model to consider 10 blocks of convolution, batch normalization, and activation layer. The learned noise is flatten and submitted to a sequence of dense and dropout layers which should learn a single float score that represents the effectiveness of the ranked list provided as input. The MAP of the synthetic data is considered as the groundtruth during training.

In all the experiments, the NAdam optimizer was used with learning rate of 10^{-4} and Mean Squared Error (MSE) loss. The network was trained considering batches of size 2, where both images correspond to the same image but with different augmentations. The method is set to have a 50% probability of thresholding the pixels of the image. If the image is selected to be thresholded, a random value is picked from 100 to 255 and all the pixels above this

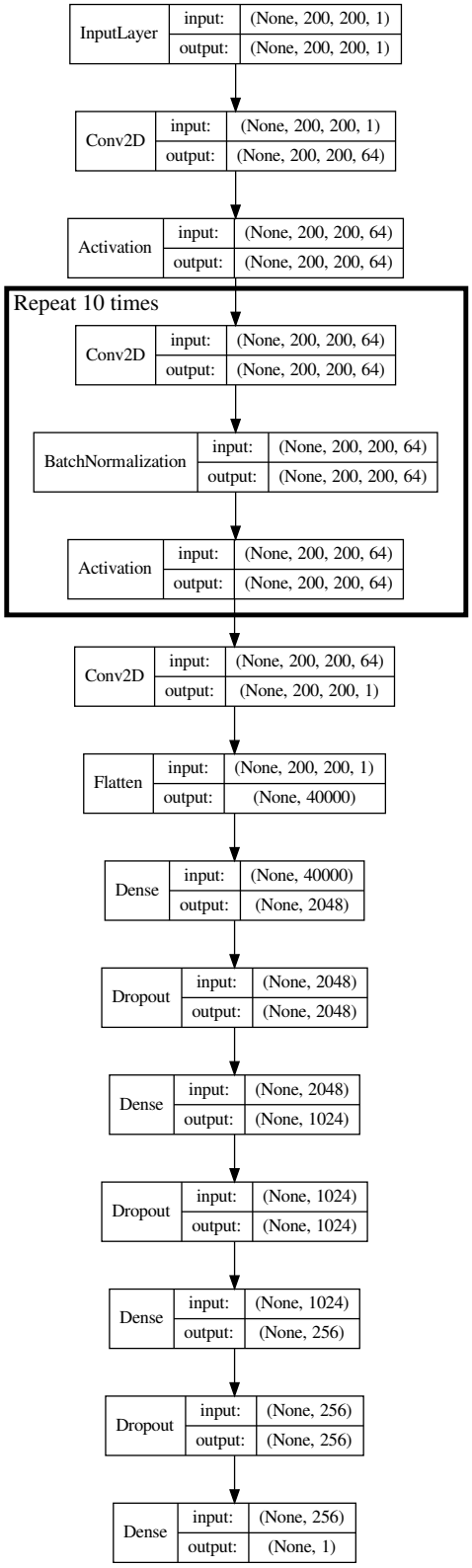


Figure 3: Proposed CNN model for effectiveness prediction.

value are set to 255. This is done with the objective of improving the network generalization during the learning process.

4 EXPERIMENTAL EVALUATION

This section describes the experiments conducted and their results. We also present comparisons with baselines in different datasets.

4.1 Experimental Settings

The experiments considered 5 different datasets with sizes ranging from 1,336 to 36,411 images:

- **OxfordFlowers-17 [36]**: 1,336 images of 17 different species of flowers (80 images per class);
- **MPEG-7 [23]**: 1,400 diverse shape images of animals, fruits, and other objects (20 images per class);
- **Brodatz [2]**: 1,776 texture images of 111 distinct classes;
- **Market1501 [66]**: a popular re-ID dataset composed of 32,217 images of 1,501 individuals (750 identities for training and 751 for testing), 3368 images of the total are considered as query images;
- **DukeMTMC-reID [68]**: a re-ID dataset composed of 36,411 images of 1,812 people (702 identities for training, 702 for testing, and 408 distractors), 2228 images of the total are considered as query images.

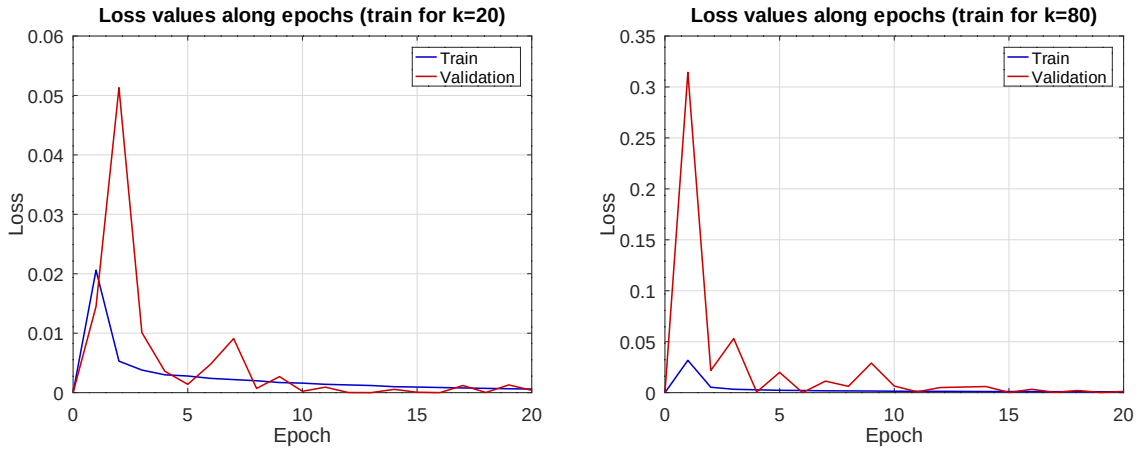
For all the datasets, the Mean Average Precision (MAP) was considered for evaluating the effectiveness. In all the cases, all images are considered as query images, except for re-ID datasets, where only query images specified by the dataset protocol were considered, as done by most of the authors in the literature. All re-ID evaluations are of single query type.

The descriptors vary in each case, according to the properties of each dataset. In total, more than 30 different descriptors are considered in this work. All the CNN extractions were performed with models trained on ImageNet [20] dataset [1], except for re-ID where the models were trained on MSMT17 [59] dataset [2].

Two different trainings were done, both of them considered artificially generated data for keeping the strategy and analysis unsupervised. The only required parameter is the size of the virtual classes (k), which impacts the images generated for training. While the first training considered $k = 20$, the second used $k = 80$. The artificial dataset contains 1,400 and 1,360 images for the trainings with $k = 20$ and $k = 80$, respectively. From the total of synthetic images, 200 of them were randomly taken for validation and the remaining was used for training. In both cases, 7 different artificial descriptors were generated with different levels of effectiveness. This adjust is done by restricting the intervals in the diagonal of the confusion matrix: first descriptor uses [0, 0.25], second uses [0, 0.5], third uses [0, 0.75], fourth uses [0, 1], fifth uses [0.25, 1], sixth uses [0.5, 1], and seventh uses [0.75, 1].

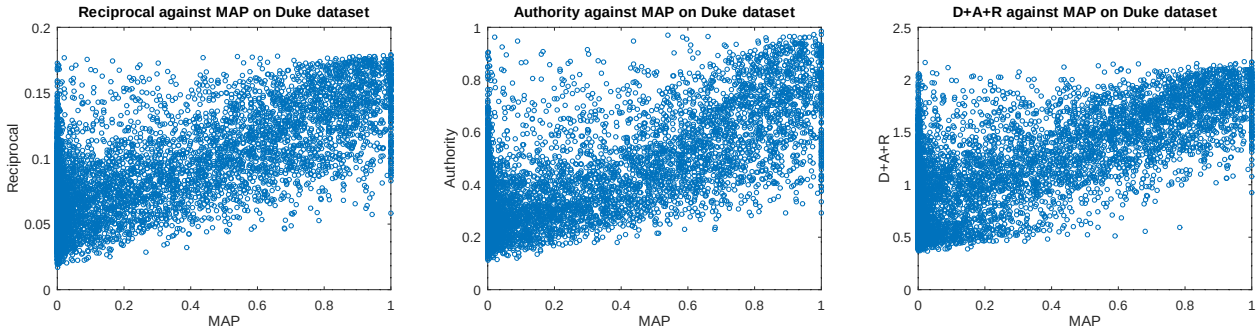
Figure 4 shows the loss values for training and validation data in both cases along 20 epochs. As can be seen, the losses decrease as the epochs increase. After 15 epochs, we can see that the model starts to decrease the train loss much slower than before, but the validation still varies. For this reason, we considered the train of 15 epochs for prediction in all the experiments to avoid overfit in the artificial data.

[1] https://github.com/Cadene/pretrained-models.pytorch
[2] https://github.com/KaiyangZhou/deep-person-reid



(a) Training for $k = 20$. **(b) Training for $k = 80$.**

Figure 4: Losses along training epochs for train and validation sets.



(a) Reciprocal (Pearson = 0.7056) **(b) Authority (Pearson = 0.7257)** **(c) DRNE+Authority+Reciprocal (Pearson = 0.7418)**

Figure 5: Correlation of MAP and effectiveness estimation measures on DukeMTMC dataset.

Our method is compared to both Authority and Reciprocal in all the cases. To keep the comparison fair, we used $k = 20$ for DRNE and the baselines in all datasets. The only exception is the Flowers dataset, where $k = 80$ was used, since it has larger classes than the others.

4.2 Experimental Results

Table 1 presents the Pearson correlation between MAP and the effectiveness estimation measures (Authority Score, Reciprocal Density, and the proposed approach) for around 30 different descriptors on Flowers dataset. To keep the comparison fair in this case, $k = 80$ was used for our approach and the baselines. The best results are highlighted in bold for each line. For the negative correlations (FOH and SCH), none of the methods were highlighted in bold. The last line of the table contains the correlation when considering all the ranked lists of all descriptors together. The original MAP of each descriptor is also presented with the objective of facilitating the analysis of the results. Notice that the best results (higher correlations) tend to be more frequent on descriptors of high effectiveness (which is the case of the CNNs). Consequently, scenarios with descriptors of low effectiveness tend to be more challenging (e.g. FOH, SCH, GIST). Besides that, the proposed approach (DRNE) achieved the best results in most of the cases, even in difficult scenarios (e.g. ACC, EHD, SPLBP). Theses cases of negative correlation occur due to the low effectiveness of such descriptors and still require more investigation.

An experiment was conducted in order to evaluate the complementary among the results provided by each effectiveness estimation measure. Table 2 shows the Pearson correlation between each pair of measures for the MPEG-7 dataset. Notice that Authority and Reciprocal are highly correlated, while our proposed approach is the less correlated with the other two, what indicates that our approach has a great potential to be combined with the others.

All the remaining datasets, where $k = 20$ was used, are presented in Table 3. Besides the individual results for each measure, combinations of measures are also presented. The combinations were done by summing the measures, also the abbreviations A, R and D were used for Authority, Reciprocal and DRNE, respectively. Notice that in most of the cases the best results correspond to our approach

Table 1: Pearson correlation between MAP and effectiveness estimation measures on Flowers dataset.

Descriptors	Original MAP	Auth.	Recipro.	DRNE (ours)
CNN-FBResNet [15]	52.56%	0.73744	0.67153	**0.79920**
CNN-ResNeXt [60]	51.91%	0.76568	0.66525	**0.79265**
CNN-ResNet [15]	51.83%	0.72981	0.63672	**0.79903**
CNN-DPNet [6]	50.93%	0.77143	0.72479	**0.79896**
CNN-Xception [7]	47.31%	0.74365	0.64060	**0.76958**
CNN-BnInception [18]	46.58%	0.57857	0.48638	**0.72061**
CNN-AlexNet [20]	46.04%	0.46586	0.35353	**0.63521**
CNN-SENet [16]	43.16%	0.58722	0.57195	**0.63076**
CNN-InceptionV4 [52]	42.35%	**0.67885**	0.58592	0.61974
CNN-InceptRN [52]	42.20%	**0.62725**	0.53364	0.55041
CNN-BnVGGNet [30]	41.87%	0.48524	0.36175	**0.63133**
CNN-NASNetLg [71]	40.74%	**0.63091**	0.55103	0.54974
CNN-VGGNet [30]	39.05%	0.50498	0.32844	**0.63850**
SIFT [31]	28.47%	0.34815	0.31624	**0.48026**
BIC [51]	25.56%	0.21481	0.16794	**0.36447**
SPJCD [32, 62]	22.56%	0.27962	0.24767	**0.33553**
SPCEDD [4, 32]	21.94%	0.31110	0.26055	**0.34731**
COMO [57]	21.83%	0.10506	0.08213	**0.25892**
SPFCTH [5, 32]	21.73%	0.19618	0.18878	**0.26632**
JCD [62]	20.89%	0.15319	0.11306	**0.24018**
FCTH [5]	20.56%	0.18428	0.13488	**0.23862**
CEDD [4]	20.48%	0.13077	0.10192	**0.20104**
SPACC [17, 32]	19.20%	0.07436	0.03312	**0.20229**
ACC [17]	18.99%	0.03264	0.02153	**0.28373**
CLD [8]	18.54%	0.32734	0.25345	**0.34693**
PHOG [12, 32]	14.74%	0.33586	0.33548	**0.37418**
SCH [8]	13.43%	-0.21997	-0.20886	-0.13598
EHD [34]	12.46%	0.03510	0.06457	**0.20214**
FOH [32, 56]	11.42%	-0.06418	-0.06645	-0.03603
SPLBP [32, 37]	10.92%	0.06942	0.07869	**0.14425**
LBP [37]	10.34%	0.01482	0.02083	**0.07323**
SCD [8]	10.25%	**0.25619**	0.10035	0.05702
GIST [38]	9.82%	-0.01581	0.02297	**0.02691**
All Descriptors	—	0.39789	0.31277	**0.42907**

Table 2: Pearson correlation between estimation measures for all descriptors of MPEG-7 dataset.

	Authority	Reciprocal	DRNE (ours)
Authority	1.0000	0.96928	0.86480
Reciprocal	0.96928	1.0000	0.87641
DRNE (ours)	0.86480	0.87641	1.0000

or a combination that involves our approach. While most of the datasets consider classes of same size, this does not occur for the re-ID datasets (Market and Duke). Even with this challenge, the results are very promising. The combination of the three measures achieved up to 0.74 of Pearson correlation in the Duke dataset, which is very significant considering that no labels were used.

Figure 5 presents a graph where each dot corresponds to a ranked list of the DukeMTMC dataset. The dots are plotted according to the value presented by the effectiveness estimation (that uses no labels) and the MAP (which uses labels). As can be seen, the results provided by the combination of the three measures present a more linear shape, and consequently a higher Pearson correlation as well.

Two visual query examples are presented on Figure 6 with the DRNE score obtained for each of them. The query image is presented in green borders and the incorrect results in red borders. Notice that DRNE attributed a lower score for the ranked list which presented wrong images and a higher score (very close to 1) for the one without errors.

Regarding execution time, the prediction time is 9.2909 ± 9.38663 milliseconds considering the mean and standard deviation for 44,880 different ranked lists. A training of 9,600 images takes about 25 minutes to run for each epoch on a NVIDIA RTX 2080 GPU. For a training of 20 epochs, it is required around 8 hours in total.

5 CONCLUSION

This work proposed a variant of the DnCNN network for effectiveness estimation on information retrieval tasks. The model was trained in a self-supervised fashion considering artificially generated data and evaluated on 5 different image retrieval datasets, including generic scenarios and re-ID ones. The measure was compared to two different effectiveness estimation measures commonly used for ranked list data in the literature (Authority Score and Reciprocal Density). The results revealed that our method provided competitive results in most of the cases. The best results could be obtained by combining the three measures.

Among the future works, we intend to investigate different variations of how to generate artificial data and possible other augmentations to train the network model and improve its generalization. Besides that, we also intend to investigate strategies to automatically define the parameter k.

ACKNOWLEDGMENTS

The authors are grateful to São Paulo Research Foundation - FAPESP (grants #2017/25908-6, #2018/15597-6, and #2020/11366-0), Brazilian National Council for Scientific and Technological Development - CNPq (grant #309439/2020-5), Petrobras (grant #2017/00285-6), and Microsoft Research for financial support.

REFERENCES

[1] Nafiz Arica and Fatos T. Yarman Vural. 2003. BAS: a perceptual shape descriptor based on the beam angle statistics. *Pattern Recognition Letters* 24, 9-10 (2003), 1627–1639.

[2] Phil Brodatz. 1966. *Textures: A Photographic Album for Artists and Designers.* Dover.

[3] Xiaobin Chang, Timothy M. Hospedales, and Tao Xiang. 2018. Multi-Level Factorisation Net for Person Re-Identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

[4] Savvas A. Chatzichristofis and Yiannis S. Boutalis. 2008. CEDD: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In *Proceedings of the 6th international conference on Computer vision systems (ICVS'08).* 312–322.

[5] Savvas A. Chatzichristofis and Yiannis S. Boutalis. 2008. FCTH: Fuzzy Color and Texture Histogram - A Low Level Feature for Accurate Image Retrieval. In *WIAMIS.* 191–196.

[6] Yunpeng Chen, Jianan Li, Huaxin Xiao, Xiaojie Jin, Shuicheng Yan, and Jiashi Feng. 2017. Dual Path Networks. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4467–4475.

[7] F. Chollet. 2017. Xception: Deep Learning with Depthwise Separable Convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 1800–1807.

[8] Leszek Cieplinski. 2001. MPEG-7 Color Descriptors and Their Applications. In *Computer Analysis of Images and Patterns*, Władysław Skarbek (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 11–20.

[9] Steve Cronen-Townsend, Yun Zhou, and W. Bruce Croft. 2002. Predicting Query Performance. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '02).* 299–306.

Table 3: Pearson correlation between MAP and effectiveness estimation measures on datasets considering train with $k = 20$.

Dataset	Descriptors	Original MAP	Auth.	Recipro.	DRNE	A+R	D+A	D+R	D+A+R
MPEG-7	**AIR** [14]	89.39%	0.76392	0.75069	0.87705	0.76419	0.86921	**0.88494**	0.86386
	ASC [28]	85.28%	0.76594	0.81430	0.74678	0.77823	0.79525	0.78578	**0.80045**
	IDSC [27]	81.70%	0.77826	**0.80911**	0.74767	0.78716	0.79826	0.78162	0.80235
	CFD [41]	80.71%	0.79817	0.83621	0.82587	0.80758	0.84616	0.84731	**0.84769**
	BAS [1]	71.52%	0.79029	**0.84011**	0.79698	0.80281	0.81826	0.82081	0.82334
	SS [10]	37.67%	0.78474	0.81322	0.84026	0.79460	0.83954	**0.84605**	0.84076
	All Descriptors	—	0.85355	0.88229	0.84607	0.86131	0.88019	0.86754	**0.88336**
Brodatz	**LAS** [53]	75.15%	0.64725	0.63484	0.69576	0.65333	0.70116	**0.70457**	0.70007
	CCOM [19]	57.57%	0.63535	0.60433	0.65631	0.63799	**0.67730**	0.66598	0.67608
	LBP [37]	48.40%	0.49609	0.42214	0.49984	0.49278	**0.52400**	0.50540	0.5199
	All Descriptors	—	0.57502	0.54266	0.59152	0.57759	**0.61023**	0.60107	0.60917
Market	**CNN-OSNET-AIN** [69]	43.30%	0.65170	0.60202	0.63451	0.64876	**0.66854**	0.64148	0.66665
	CNN-HACNN [25]	23.30%	0.52763	0.48562	0.52421	0.52611	**0.54461**	0.52853	0.54371
	CNN-ResNet [15]	22.82%	0.60783	0.55807	0.60246	0.60517	**0.62471**	0.60710	0.62385
	CNN-MLFN [3]	21.98%	0.57916	0.53273	0.55649	0.57662	**0.58287**	0.56158	0.58243
	BOVW [66]	13.34%	0.39235	0.31576	0.38518	0.3832	**0.40171**	0.38429	0.3983
	WHOS [29]	6.23%	0.13383	0.14891	0.22140	0.13952	0.20209	**0.21919**	0.2005
	All Descriptors	—	0.61279	0.53534	0.57867	0.60599	**0.61197**	0.58337	0.61038
Duke	**CNN-OSNET-AIN** [69]	52.69%	0.64525	0.64349	0.64988	0.64861	0.67623	0.66199	**0.67631**
	CNN-ResNet [15]	32.00%	0.69101	0.67709	0.67628	0.69335	0.70446	0.68566	**0.70552**
	WHOS [29]	2.65%	0.00572	0.03644	**0.11433**	0.01163	0.07868	0.10800	0.07615
	All Descriptors	—	0.72574	0.70560	0.71234	0.72660	0.74163	0.72138	**0.74189**

(a) Ranked list with DRNE score of 0.5153

(b) Ranked list with DRNE score of 0.9650

Figure 6: Two examples of ranked lists (good and bad queries) for Duke dataset and OSNET-AIN descriptor.

[10] Ricardo da S. Torres and Alexandre X. Falcão. 2007. Contour Salience Descriptors for Effective Image Retrieval and Analysis. *Image and Vision Computing* 25, 1 (2007), 3–13.

[11] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. 2007. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Transactions on Image Processing* 16, 8 (2007), 2080–2095.

[12] N. Dalal and B. Triggs. 2005. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1. 886–893 vol. 1.

[13] Shiv Ram Dubey. 2021. A Decade Survey of Content Based Image Retrieval using Deep Learning. *IEEE Transactions on Circuits and Systems for Video Technology* (2021), 1–1.

[14] Raghuraman Gopalan, Pavan Turaga, and Rama Chellappa. 2010. Articulation-invariant representation of non-planar shapes. In *ECCV*, Vol. 3. 286–299.

[15] K. He, X. Zhang, S. Ren, and J. Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778.

[16] Jie Hu, Li Shen, and Gang Sun. 2018. Squeeze-and-Excitation Networks. In *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[17] Jing Huang, S. Ravi Kumar, Mandar Mitra, Wei-Jing Zhu, and Ramin Zabih. 1997. Image Indexing Using Color Correlograms. In *CVPR*. 762–768.

[18] Sergey Ioffe and Christian Szegedy. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37 (ICML'15)*. JMLR.org, 448–456.

[19] Vassili Kovalev and Stephan Volmer. 1998. Color Co-occurence Descriptors for Querying-by-Example. In *International Conference on Multimedia Modeling (ICMM)*. 32.

[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12)*. Curran Associates Inc., USA, 1097–1105.

[21] A. Krull, T. Buchholz, and F. Jug. 2019. Noise2Void - Learning Denoising From Single Noisy Images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2124–2132.

[22] Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. 2019. High-Quality Self-Supervised Deep Image Denoising. In *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.). Curran Associates, Inc., 6970–6980.

[23] Longin Jan Latecki, Rolf Lakamper, and Ulrich Eckhardt. 2000. Shape Descriptors for Non-rigid Shapes with a Single Closed Contour. In *CVPR*. 424–429.

[24] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. 2018. Noise2Noise: Learning Image Restoration without Clean Data *(Proceedings of Machine Learning Research, Vol. 80)*, Jennifer Dy and Andreas Krause (Eds.). PMLR, Stockholmsmässan, Stockholm Sweden, 2965–2974.

[25] Wei Li, Xiatian Zhu, and Shaogang Gong. 2018. Harmonious Attention Network for Person Re-Identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[26] Yangxi Li, Bo Geng, Linjun Yang, Chao Xu, and Wei Bian. 2012. Query difficulty estimation for image retrieval. *Neurocomputing* 95 (2012), 48–53.

[27] Haibin Ling and David W. Jacobs. 2007. Shape Classification Using the Inner-Distance. *IEEE TPAMI* 29, 2 (2007), 286–299.

[28] Haibin Ling, Xingwei Yang, and Longin Jan Latecki. 2010. Balancing Deformability and Discriminability for Shape Matching. In *ECCV*, Vol. 3. 411–424.

[29] G. Lisanti, I. Masi, A. D. Bagdanov, and A. D. Bimbo. 2015. Person Re-Identification by Iterative Re-Weighted Sparse Ranking. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, 8 (2015), 1629–1642.

[30] S. Liu and W. Deng. 2015. Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*. 730–734.

[31] D.G. Lowe. 1999. Object recognition from local scale-invariant features. In *ICCV*. 1150–1157.

[32] Mathias Lux. 2011. Content Based Image Retrieval with LIRe. In *Proceedings of the 19th ACM International Conference on Multimedia (MM '11)*. ACM, New York, NY, USA, 735–738.

[33] L. V. D. Maaten and Geoffrey E. Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9 (2008), 2579–2605.

[34] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada. 2001. Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology* 11, 6 (Jun 2001), 703–715.

[35] N. Moran, D. Schmidt, Y. Zhong, and P. Coady. 2020. Noisier2Noise: Learning to Denoise From Unpaired Noisy Data. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 12061–12069.

[36] M-E. Nilsback and A. Zisserman. 2006. A Visual Vocabulary for Flower Classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2. 1447–1454.

[37] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. 2002. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *PAMI* 24, 7 (2002), 971–987.

[38] Aude Oliva and Antonio Torralba. 2001. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *IJCV* 42, 3 (May 2001), 145–175.

[39] A. Oliveira and A. Rocha. 2020. Contextual Features and Sequence Labeling Techniques for Relevance Prediction in Retrieval. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. 305–310.

[40] D. C. G. Pedronette and R. d. S. Torres. 2015. Unsupervised Effectiveness Estimation for Image Retrieval Using Reciprocal Rank Information. In *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*. 321–328.

[41] Daniel Carlos Guimarães Pedronette and Ricardo da S. Torres. 2010. Shape Retrieval using Contour Features and Distance Optmization. In *VISAPP*, Vol. 1. 197 – 202.

[42] Daniel Carlos Guimarães Pedronette and Ricardo da Silva Torres. 2015. Unsupervised Effectiveness Estimation for Image Retrieval Using Reciprocal Rank Information. In *28th SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI*. 321–328.

[43] Daniel Carlos Guimarães Pedronette, Otávio A.B. Penatti, and Ricardo da S. Torres. 2014. Unsupervised manifold learning using Reciprocal kNN Graphs in image re-ranking and rank aggregation tasks. *Image and Vision Computing* 32, 2 (2014), 120 – 130.

[44] Daniel Carlos Guimarães Pedronette and Ricardo da S. Torres. 2012. Exploiting contextual information for image re-ranking and rank aggregation. *International Journal of Multimedia Information Retrieval* 1, 2 (Jul 2012), 115–128.

[45] Luca Piras and Giorgio Giacinto. 2017. Information fusion in content based image retrieval: A comprehensive overview. *Information Fusion* 37, Supplement C (2017), 50 – 60.

[46] João Gabriel Camacho Presotto, Lucas Pascotti Valem, and Daniel Carlos Guimarães Pedronette. 2019. Unsupervised Effectiveness Estimation Through Intersection of Ranking References. In *Computer Analysis of Images and Patterns - CAIP 2019*, Vol. 11679. 231–244.

[47] Y. Quan, M. Chen, T. Pang, and H. Ji. 2020. Self2Self With Dropout: Learning Self-Supervised Denoising From Single Image. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1887–1895.

[48] Anuraganand Sharma and Dinesh Kumar. 2020. Non-image Data Classification with Convolutional Neural Networks. (07 2020).

[49] Alok Sharma, Edwin Vans, Daichi Shigemizu, Keith A. Boroevich, and Tatsuhiko Tsunoda. 2019. DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture. *Scientific Reports* 9, 1 (Aug 2019), 11399.

[50] Shakarim Soltanayev and Se Young Chun. 2018. Training deep learning based denoisers without ground truth data. In *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Eds.). Curran Associates, Inc., 3257–3267.

[51] Renato O. Stehling, Mario A. Nascimento, and Alexandre X. Falcão. 2002. A compact and efficient image retrieval approach based on border/interior pixel classification. In *CIKM*. 102–109.

[52] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. 2017. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. (2017).

[53] Bo Tao and Bradley W. Dickinson. 2000. Texture Recognition and Image Retrieval Using Gradient Indexing. *JVCIR* 11, 3 (2000), 327–342.

[54] Ricardo Da Silva Torres and Alexandre Xavier Falcão. 2006. Content-Based Image Retrieval: Theory and Applications. *Revista de Informática Teórica e Aplicada* 13 (2006), 161–185.

[55] Lucas Pascotti Valem and Daniel Carlos Guimarães Pedronette. 2020. Unsupervised selective rank fusion for image retrieval tasks. *Neurocomputing* 377 (2020), 182–199.

[56] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. 2010. Evaluating Color Descriptors for Object and Scene Recognition. *PAMI* 32, 9 (2010), 1582–1596.

[57] Sotiris A. Vassou, Nektarios Anagnostopoulos, Angelos Amanatiadis, Klitos Christodoulou, and Savvas A. Chatzichristofis. 2017. CoMo: A Compact Composite Moment-Based Descriptor for Image Retrieval. In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing (CBMI '17)*. ACM, New York, NY, USA, 30:1–30:5.

[58] Charles Webber, Cornel Ioana, and Nobert Marwan. 2016. *Recurrence Plots and Their Quantifications: Expanding Horizons: Proceedings of the 6th International Symposium on Recurrence Plots, Grenoble, France, 17-19 June 2015.*

[59] L. Wei, S. Zhang, W. Gao, and Q. Tian. 2018. Person Transfer GAN to Bridge Domain Gap for Person Re-identification. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 79–88.

[60] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He. 2017. Aggregated Residual Transformations for Deep Neural Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 5987–5995.

[61] Xing Xing, Yi Zhang, and Mei Han. 2010. Query Difficulty Prediction for Contextual Image Retrieval. In *European Conference on IR Research, ECIR*, Vol. 5993. 581–585.

[62] K. Zagoris, S.A. Chatzichristofis, N. Papamarkos, and Y.S. Boutalis. 2010. Automatic Image Annotation and Retrieval Using the Joint Composite Descriptor. In *14th Panhellenic Conference on Informatics (PCI)*. 143–147.

[63] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. 2020. Plug-and-Play Image Restoration with Deep Denoiser Prior. *arXiv preprint* (2020).

[64] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. 2017. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing* 26, 7 (2017), 3142–3155.

[65] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. 2018. Residual Dense Network for Image Super-Resolution. In *CVPR*.

[66] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. 2015. Scalable Person Re-identification: A Benchmark. In *2015 IEEE International Conference on Computer Vision (ICCV)*. 1116–1124.

[67] L. Zheng, Y. Yang, and Q. Tian. 2018. SIFT Meets CNN: A Decade Survey of Instance Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 5 (2018), 1224–1244.

[68] Zhedong Zheng, Liang Zheng, and Yi Yang. 2017. Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro. In *Proceedings of the IEEE International Conference on Computer Vision*. 3754–3762.

[69] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. 2021. Learning Generalisable Omni-Scale Representations for Person Re-Identification. *TPAMI* (2021).

[70] Wengang Zhou, Houqiang Li, and Qi Tian. 2017. Recent Advance in Content-based Image Retrieval: A Literature Survey. arXiv:1706.06064 [cs.MM]

[71] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. 2017. Learning Transferable Architectures for Scalable Image Recognition. *CoRR* abs/1707.07012 (2017).