Can I make a wish?: a competition on detecting meteors in images

A. C. Lorena[1], D. S. Kaster[2], R. Cerri[3], E. R. Faria[4], V. V. de Melo[5]

[1] Instituto Tecnológico de Aeronáutica, Divisão de Ciência de Computação
aclorena@ita.br
[2] Universidade Estadual de Londrina, Departamento de Computação
dskaster@uel.br
[3] Universidade Federal de São Carlos, Departamento de Computação
cerri@ufscar.br
[4] Universidade Federal de Uberlândia, Faculdade de Computação
elaine@ufu.br
[5] Universidade Federal de São Paulo, Instituto de Ciência e Tecnologia
vinicius.melo@unifesp.br

Abstract. Promoting competitions has become a path towards attracting people's interest into diverse areas. Many international conferences have sessions dedicated to one or more competitions, in which participants are challenged by real problems for which advanced solutions are needed. This paper describes the first Brazilian competition on Knowledge Discovery in Databases (KDD-BR), which was part of three main events of the Brazilian Computer Society dedicated to Artificial Intelligence, Databases and Data Mining. In this first edition the participants were supposed to detect meteors, popularly known as shooting stars, in regions of interest of images collected from a monitoring station located at São José dos Campos, Brazil. The data set assembled is detailed, which may be of interest for future benchmark studies using such data. The competition results, contributions and limitations are also discussed, providing a guide for future editions.

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database Applications; I.2.6 [**Artificial Intelligence**]: Learning

Keywords: competition, data mining, machine learning

1. INTRODUCTION

Numerous well-known international conferences and symposia have the practice of promoting competitions as one of their activities. These competitions have become routine in events such as: *Neural Information Processing Systems* (NIPS), *IEEE World Congress on Computational Intelligence* (WCCI), *International Conference on Machine Learning* (ICML) and the *International Conference on Knowledge Discovery and Data Mining* (KDD). Registered applicants must offer solutions to challenging problems from a variety of domains, such as text classification, handwriting digit recognition, marketing, and others. One of the pioneering events to promote a Machine Learning (ML) and Data Mining (DM) competition was KDD, which helds annually the KDD-Cup competition since 1997 [Rosset and Inger 2000].

The importance of such competitions is multiple: to promote the formulation of new DM and ML techniques and solutions to challenging problems; to motivate more public to participate in the underlying event; to introduce new application domains suited for DM and ML solution; among others. The used data sets are usually made publicly available, enabling any interested party, whether academic or not, to participate. According to Isabelle Guyon, a major promoter of international competitions in ML, “the challenges launched each year have allowed us to cross the frontiers of ML research” [Guyon et al. 2011]. Since the beginning of the 2010’s, some platforms which host data science and ML competitions have also been launched. One of the most popular representatives is *Kaggle* [Carpenter 2011], which has become a standard platform in which large companies launch challenges that require ML solutions. The authors of the best solutions can be rewarded in a variety of ways, from financially through job offerings.

In Brazil, competitions such as *Robocup*, in the area of robotics, are responsible for attracting a large audience and arouse the interest from the general public towards this area. In 2017, a first Brazilian competition on KDD (KDD-BR) was launched as part of the joint activities of the *Brazilian Conference on Intelligent Systems* (BRACIS), *Brazilian Symposium on Databases* (SBBD) and *Symposium on Knowledge Discovery, Mining and Learning* (KDMiLe) events. In this first edition, the participants were challenged to create an automatic algorithm able to predict whether a given region of interest in a night sky image contains or not a meteor, popularly known as a shooting star.

Monitoring meteors is of interest of some major aerospace agencies, such as NASA, which finances the Center for Near-Earth Objects Studies (CNEOS) in the California Institute of Technology [Chodas 2015]. Such observations may support defense mechanisms against possible harmful impacts on the Earth or, more commonly, to identify pieces which may be collected for chemical studies. In Brazil there are also some citizen science initiatives dedicated to the monitoring of meteors crossing the southern skies, such as EXOSS (Exploring the Southern Sky)[1] and BRAMON (Brazilian Meteor Observation Network)[2]. The data set collected for the competition is composed of images from a monitoring station of the EXOSS Citizen Science project, located at the Observatory of Astronomy and Space Physics from University of Vale do Paraíba (UNIVAP), São José dos Campos, Brazil.

The competition was launched on July, 1st, 2017 in the Kaggle in class platform. A total of 28 teams joined the competition. The participants were mostly from the southern region of Brazil, although there have been some submissions from other countries too. The top three teams were invited to present their solutions at the joint 2017 BRACIS-SBBD-KDMiLe and the actual final positions were revealed at the conference dinner, on October 4th. This first attempt to promote a KDD competition in Brazil can be considered successful and was able to attract researchers from both academy and industry for the event. This paper describes the data set made publicly available (Section 2), presents the competition configuration (Section 3), results and main statistics (Section 4), and also discusses some limitations of this first competition which can be addressed in future editions (Section 5).

2. THE DATA SET

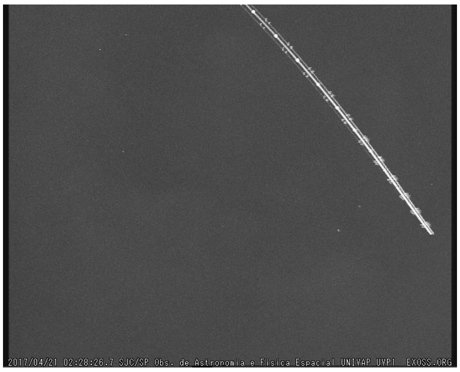
EXOSS is a Brazilian non-profit organization whose objective is to monitor meteors that cross the southern skies with a low cost system. Any citizen can apply for participation and build his/her own monitoring station. There are currently about 50 active EXOSS monitoring stations at various locations of the Brazilian territory[3]. It is a citizen science project in which data records of meteors captured by each station are gathered in a common repository after confirmation. The simultaneous capture of a same meteor by multiple stations can allow to determine its trajectory and possible impact point (most of the meteors actually get destroyed in the Earth’s atmosphere). The University of Vale do Paraíba (UNIVAP) is a partner institution of EXOSS and has a monitoring station located at its Observatory of Astronomy and Space Physics. The data set assembled for this competition is composed of images from one of the cameras of the UNIVAP monitoring station.

[1]`http://press.exoss.org/`
[2]`http://www.bramonmeteor.org/bramon/`
[3]`http://press.exoss.org/associados/estacoes-associadas-a-exoss/`

The UNIVAP monitoring station consists of a low-cost video surveillance camera with a dedicated motion capture software. This software, named *UFO Capture*, records any moving object detected by the camera, which was programmed to work during the nighttime. These objects can be either meteors or non-meteors, e.g., birds, insects, planes, lightning, and rain drops. The weather conditions and the sky configuration vary over the nights, with the presence/absence of stars, the moon or even clouds. The camera can also make noisy recordings, in which no object is indeed detected. The interest is to accurately identify those images which contain meteor records. Some examples of captured images are presented in Figure 1.

(a) Example of meteor.

(b) Example of meteor on a cloudy day.



(c) Example of non-meteor (possibly a plane).

(d) Example of non-meteor (possibly a bird/insect).

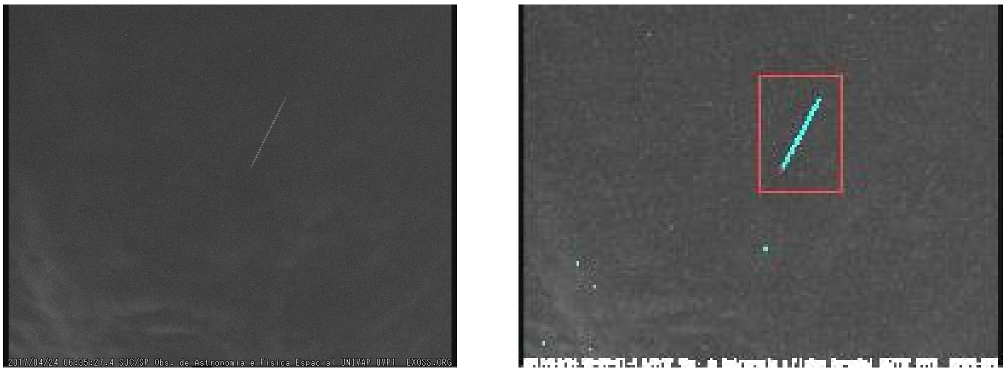
Fig. 1. Examples of captured images.

The *UFO Capture* software is distributed by the SonataCo[4] network, a Japanese initiative on low-cost meteor monitoring [Jenniskens 2017]. According to its manual, it is a motion capture software which starts recording from a few seconds before the action is recognized to a few seconds after the action finishes. The same network also distributes the *UFOAnalyzer* and *UFOOrbit* tools, which can be used in the analysis of the captured images. *UFOAnalyzer* calculates the direction and elevation of the event that is recorded, and allows to roughly confirm whether the moving object is a meteor, since some records may not correspond to a valid meteor trajectory. *UFOOrbit* is used to get the orbit of a same meteor observed by more than two different locations (monitoring stations). In this case, the object can be indeed confirmed as a meteor, which was visible at multiple sites.

[4]`http://sonotaco.com/soft/e_index.html`

The competition images were collected during the months of April and May of 2017 and were categorized by Jennifer Nielsen, Aeronautics Engineering student at UNIVAP, under the supervision of Dr Irapuan Rodrigues, Physics and Astronomy Professor at UNIVAP. Periodically, this student scans the videos and images from the UNIVAP monitoring station and deletes all non-meteors files. This competition aims to built an automate system to support this filtering, by classifying the recorded images into two classes: meteor vs non-meteor. This shall reduce the overhead of manually filtering the non-meteor cases from the daily repository formed locally at UNIVAP.

The *UFO Capture* software stores five files per recording: (i) a movie in the AVI format; (ii) an XML file with profile information for *UFO Analyzer*; (iii) a bitmap file with mask and average brightness information for *UFOAnalyzer*; (iv) a JPEG file containing a peak hold or snapshot still image of the captured event; and (v) a thumbnail JPEG image, in which a region of interest where the moving object was detected is also highlighted in a rectangle. The regions of interest in the snapshot images were used in the competition. Therefore, first the highlighted regions of the thumbnail images (as shown in Figure 2b) were identified in the corresponding snapshot images (shown in Figure 2a), which were cropped. We opted to use the snapshot images, which had a better resolution.



(a) Snapshot of a meteor.   (b) Thumbnail image of the same meteor.

Fig. 2. Examples of two JPEG images stored by *UFO Capture* per recording.

A total of 122 images were captured and labeled: 41 meteors and 81 non-meteors. For each image, a large set of characteristics were extracted by various image processing algorithms from the JFeatureLib library[5] using a workflow-based image retrieval distributed architecture [Milano-Oliveira and Kaster 2017], as shown in Table I. A total number of 3, 451 features were extracted. The idea was to build a data set with diverse information about the images, so that competitors could opt to use all or part of them in their automatic system. We opted to extract those features instead of distributing the images directly for avoiding the competition to be biased towards image processing solutions only. Indeed, the distributed data set has challenging characteristics for ML and DM: it has a high dimensionality and a low number of examples; it is noisy; and it is slightly imbalanced with a ratio (ratio of the number of examples in the majority class to the number of examples in the minority class) of 1.97, which can be considered moderate [Fernández et al. 2008].

3. KAGGLE CONFIGURATION

The competition was hosted in the *Kaggle* platform[6] and launched on July 1st, 2017. In particular, we used *Kaggle in class*, which is designed for hosting academic ML competitions at no cost.

[5]`https://github.com/locked-fg/JFeatureLib`
[6]https://www.kaggle.com/c/can-i-make-a-wish-detecting-shooting-stars

Table I. Feature sets extracted from the images.

Feature set	**# features**	**Reference**
Auto Color Correlogram	768	[Huang et al. 1997]
CEDD	144	[Chatzichristofis and Boutalis 2008a]
Color Histogram	64	[Novak and Shafer 1992]
FCTH	192	[Chatzichristofis and Boutalis 2008b]
Fuzzy Histogram	125	[Han and Ma 2002]
Fuzzy Opponent Histogram	576	[Van De Sande et al. 2010]
Gabor	60	[Fogel and Sagi 1989]
Haralick	14	[Haralick et al. 1973]
Histogram	256	[Scott 2010]
JCD	168	[Zagoris et al. 2010]
Jpeg Coefficient Histogram	192	[Sikora 2001]
Luminance Layout	64	[Sikora 2001]
MPEG7 Color Layout	33	[Sikora 2001]
MPEG7 Edge Histogram	80	[Sikora 2001]
Mean Intensity Local Binary Patterns	256	[Ojala et al. 1994]
Mean Patch Intensity Histogram	256	[Taylor and Drummond 2011]
Moments	4	[Abo-Zaid et al. 1988]
Opponent Histogram	64	[van de Sande et al. 2004]
PHOG	40	[Bosch et al. 2007]
Reference Color Similarity	77	[Kriegel et al. 2011]
Tamura	18	[Tamura et al. 1978]

The data set was randomly divided into a training and a testing sets. The test set was further randomly divided into two halves by the *Kaggle* platform: public and private. During any *Kaggle* competition, a leaderboard is built based on the performance achieved on the public test data partition. At the end of the competition, the performance on the private data is also revealed and the competitors are ranked accordingly. An overfitted model can present high-quality results on the public test set, achieving top rank positions in the public leaderboard but a much lower rank on the private leaderboard. For that reason, competitors must do their best at avoiding overfitting. This effect was observed in this competition, in which some competitors with low performance on the private data were top-ranked in the public data and vice versa.

Kaggle supports a wide range of evaluation measures. Initially, the meteor competition was configured towards maximizing the AUC (Area Under the ROC Curve) measure, which revealed to be non-competitive on our data set. For this reason, about one week later the log-loss was adopted instead, and it had to be minimized. The final results of the competition were based on the solutions posted until September 18th, 2017. After the deadline, the system was still open for submissions, but they were not taken into account for computing the final competition results. This was done so that the final ranking could be revealed during the conference only. The top three teams were invited to present their solutions at a competition award session, on October 3rd, 2017, but the actual final positions of the ranking were disclosed during the conference dinner on October 4th, 2017.

3.1 Rules

Mostly, the rules of the competition were kept the standard suggested in the *Kaggle in class* platform: (i) the participants were allowed to form teams; (ii) team mergers were allowed only for teams containing one member each; and (iii) each participant could submit a limit of two solutions per day and could opt for two final submissions for judging.

The submission files had to be formatted with three columns: Id, Prob1 and Prob2. Id gives the identification of the test image. The values Prob1 and Prob2 correspond to the predicted probabilities towards class non-meteor and meteor, respectively.

3.2 Data sets

The training set available for the competitors was composed of 80 labeled instances of which 54 were *non-meteors* and 26 were *meteors*. The test set had 42 unlabeled instances with 27 *non-meteors* and 15 *meteors*. Meteors were labeled as 1 and non-meteors as 0. We provided two types of data sets. The training and test sets contained all features extracted from the images, while a *zip* file named *DatasetPerFeature.zip* contained the training and testing partitions separated according to each type of characteristic extracted from the images as presented in Table I.

3.3 Evaluation

The evaluation metric for this competition was the log-loss, which evaluates the accuracy of a classifier by penalizing false classifications. Thus, minimizing the log-loss is similar to maximizing the classifier's accuracy. As a result, a perfect classifier has a log-loss of zero, while the remaining classifiers have progressively larger values. Log-loss is based on uncertainty; thus, the classifier predictions must be probabilities. For a binary classification problem, the log-loss expression is: $-\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1-y_i)\log(1-p_i)]$, where N is the data set size, y_i is the ground truth (correct class) of instance i, p_i is the predicted probability for instance i, and log is the natural logarithm. An important property of that metric is that it penalizes heavily when the model makes incorrect predictions. For instance, if $y_i = 1$ and $p_i = 0.5$, then log-loss ≈ 0.69. On the other hand, a confident mistake of $p_i = 0.001$ results in log-loss ≈ 6.9.

4. COMPETITION RESULTS

The competition attracted 32 participants, which were organized in 28 teams. An average of 10.78 submissions were done per team, with a standard deviation of 10.18. Whilst there were unique submissions from some participants, a team submitted up to 39 solutions during the competition. The participants were mostly from the Southern region of Brazil (São Paulo, Minas Gerais and Rio de Janeiro states), but there were also some participants from other countries, namely Peru, South Africa, United States and India.

The log-loss results achieved in the competition are shown in Table II. This table shows the average, standard deviation, minimum, and maximum log-loss performance achieved by the competitors on the public and private test data. The average performances in both test data partitions are quite similar, although the average performance on the public data was a little worse. In both partitions, it was possible to attain a null log-loss, as evidenced in the Minimum column.

Table II. Competition results (log-loss).

Test set	**Average**	**Standard-deviation**	**Minimum**	**Maximum**
Public	0.555	0.446	0.000	1.501
Private	0.459	0.483	0.000	2.502

The third top-ranked solution was proposed by Victor Almeida (public log-loss of 0.597 and private log-loss of 0.117), from the Federal Fluminense University and Petrobras, and is based on off-the-shelf algorithms. The second place was achieved by a team from the Federal University of São Carlos, composed of Renato Silva, Tiago Almeida, and Johannes Lochter (public log-loss of 0.000 and private log-loss of 0.071). The solution[7] is a stacking approach that uses a meta-classifier that is trained with the probabilities given by individual models such that each individual model is trained with the training data represented by one of the 21 feature sets available. The winning team, composed of

[7] `http:\\https://github.com/renatoms88/KDDBR`

Humberto Brandão and Hugo Pinto (public log-loss of 0.000 and private log-loss of 0.000), employed a proprietary optimization process that tries tons of mathematical expressions in order to "recreate" a function that exactly represents the problem. If the answer of the obtained function for an instance is greater than 0.5, it is classified as a meteor. Otherwise, it is a non-meteor. This solution achieved a null log-loss, however, the underlying method was not publicized.

Looking at the log information from the competition, it was possible to notice some interesting points. Firstly, the third ranked team on private data achieved a bottom public position (20th position). There were also teams which achieved a good performance on the public test set, but were in bottom positions for private data. Many teams did not submit top solutions too, since the participant can choose which submissions will be judged by the system. Curiously, two out of the three winning teams would be different if some participants were judged by other submitted solutions.

5. DISCUSSION

Based on the results and participation, the 1st KDD-BR Competition can be considered successful. Nonetheless, during the organization of the KDD-BR competition, some issues were noticed, which are here described. First, configuring the competition was not easy. To obtain a real world classification problem with a labeled data set was challenging. For this step, we counted with the help of Observatory of Astronomy and Space Physics from University of Vale do Paraíba (UNIVAP), which provided the labeled images from meteors and non-meteors. The possibility of creating an artificial data set was also considered by the organizing team at some point, but a real world data set was considered more attractive.

The second step was to extract features from the captured images, in order to build a data set to be used in the competition. A very diverse set of characteristics was extracted from the images, using multiple feature extractors. Some extractors may have produced irrelevant information to the problem, but identifying this irrelevant information has brought an additional challenge to the competition. Moreover, the organizing committee preferred to provide the feature vectors instead of the original images, because the focus of the competition was on comparing ML-DM solutions.

The third step was to configure the competition in the *Kaggle in class* platform. Around 66.6% of the data set was used as training data and the test set was split in public and private sets. At this step, the choice of the evaluation measure was an important issue to be discussed. At the beginning of the competition, the AUC measure was chosen to evaluate the developed solutions. However, most of the teams achieved 100% of AUC in the public leaderboard, which did not motivate the competitors to improve their solutions. At this time, about one week after the competition started, the evaluation measure was exchanged into the log-loss, making the competition more attractive. The final winners of the competition were the teams that achieved the best log-loss values in the private leaderboard. As previously discussed, one of the top-three teams on private data is at a bottom public position. In addition, there were teams with good public results in bottom positions for private data.

The top-three teams were invited to present their solutions on the conference. However, the solutions developed by some of the teams were not made public, since this was not a requirement of the competition. This makes it impossible to reproduce the obtained results and prevents their application to other contexts. Another issue was the size of the data set, which was too small and biased the results towards the particular data partition used for testing. In this case, a leave-one-out evaluation strategy would be more indicated. During the competition the organizers also revealed the private leaderboard publicly, omitting the five top-ranked team names. This showed to be a leak of information which could be avoided.

Acknowledgements

To Luis Fernando Milano Oliveira and Rodrigo Kuerten, from UEL, for pre-processing the images. To Jennifer Nielsen and Prof Dr Irapuan Rodrigues, from UNIVAP, for collecting the images. Finally, to the financial support of FAPESP (grant 2012/22608-8) and CNPq (grants 308858/2014-0, 305291/2017-3, and 426202/2016-3).

REFERENCES

Abo-Zaid, A., Hinton, O. R., and Horne, E. About moment normalization and complex moment descriptors. In *Pattern Recognition*. Springer, pp. 399–409, 1988.

Bosch, A., Zisserman, A., and Munoz, X. Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*. ACM, pp. 401–408, 2007.

Carpenter, J. May the best analyst win. *Science* 331 (6018): 698–699, 2011.

Chatzichristofis, S. A. and Boutalis, Y. S. Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In *Int. Conf. on Computer Vision Systems*. Springer, pp. 312–322, 2008a.

Chatzichristofis, S. A. and Boutalis, Y. S. Fcth: Fuzzy color and texture histogram-a low level feature for accurate image retrieval. In *Image Analysis for Multimedia Interactive Services, WIAMIS'08*. IEEE, pp. 191–196, 2008b.

Chodas, P. Overview of the jpl center for neo studies (cneos). In *AAS/Division for Planetary Sciences Meeting Abstracts*. Vol. 47, 2015.

Fernández, A., García, S., del Jesus, M. J., and Herrera, F. A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. *Fuzzy Sets and Systems* 159 (18): 2378–2398, 2008.

Fogel, I. and Sagi, D. Gabor filters as texture discriminator. *Biological cybernetics* 61 (2): 103–113, 1989.

Guyon, I., Cawley, G., Dror, G., and Saffari, A. Hands-on pattern recognition challenges in machine learning, volume, 2011.

Han, J. and Ma, K.-K. Fuzzy color histogram and its use in color image retrieval. *IEEE transactions on image processing* 11 (8): 944–952, 2002.

Haralick, R. M., Shanmugam, K., et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics* (6): 610–621, 1973.

Huang, J., Kumar, S. R., Mitra, M., Zhu, W.-J., and Zabih, R. Image indexing using color correlograms. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 762–768, 1997.

Jenniskens, P. Meteor showers in review. *Planetary and Space Science* vol. 143, pp. 116–124, 2017.

Kriegel, H.-P., Schubert, E., and Zimek, A. Evaluation of multiple clustering solutions. In *MultiClust@ ECML/PKDD*. pp. 55–66, 2011.

Milano-Oliveira, L. F. and Kaster, D. Defining similarity spaces for large-scale image retrieval through scientific workflows. In *Proceedings of the 21st Int. Database Engineering & Applications Symposium*. ACM, pp. 57–65, 2017.

Novak, C. L. and Shafer, S. A. Anatomy of a color histogram. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*. IEEE, pp. 599–605, 1992.

Ojala, T., Pietikainen, M., and Harwood, D. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition*. Vol. 1. IEEE, pp. 582–585, 1994.

Rosset, S. and Inger, A. Kdd-cup 99: knowledge discovery in a charitable organization's donor database. *SIGKDD Explorations* 1 (2): 85–90, 2000.

Scott, D. W. Averaged shifted histogram. *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (2): 160–164, 2010.

Sikora, T. The mpeg-7 visual standard for content description-an overview. *IEEE Transactions on circuits and systems for video technology* 11 (6): 696–702, 2001.

Tamura, H., Mori, S., and Yamawaki, T. Textural features corresponding to visual perception. *IEEE Transactions on Systems, man, and cybernetics* 8 (6): 460–473, 1978.

Taylor, S. and Drummond, T. Binary histogrammed intensity patches for efficient and robust matching. *International journal of computer vision* 94 (2): 241–265, 2011.

Van De Sande, K., Gevers, T., and Snoek, C. Evaluating color descriptors for object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence* 32 (9): 1582–1596, 2010.

van de Sande, K. E., Gevers, T., and Snoek, C. G. Evaluation of color descriptors for object and scene recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, USA (June 2008)*, 2004.

Zagoris, K., Chatzichristofis, S. A., Papamarkos, N., and Boutalis, Y. S. Automatic image annotation and retrieval using the joint composite descriptor. In *Informatics (PCI), 2010 14th Panhellenic Conference on*. IEEE, pp. 143–147, 2010.