Classification of biological cells using bio-inspired descriptors

Wafa Bel Haj Ali, Dario Giampaglia, Michel Barlaud, Paolo Piro, Richard Nock, Thierry Pourcher

Classification of biological cells using bio-inspired descriptors

Wafa Bel haj ali, Dario Giampaglia, Michel Barlaud
I3S Lab., CNRS/University of Nice-Sophia Antipolis, France
{*belhajal, giampagl, barlaud*}*@i3s.unice.fr*

Paolo Piro
Istituto Italiano di Tecnologia, Genoa, Italy
paolo.piro@iit.it

Richard Nock
CEREGMIA, University of Antilles-Guyane, Martinique, France
rnock@martinique.univ-ag.fr

Thierry Pourcher
CEA/University of Nice-Sophia Antipolis, France
pourcher@unice.fr

Abstract

This paper proposes a novel automated approach for the categorization of cells in fluorescence microscopy images. Our supervised classification method aims at recognizing patterns of unlabeled cells based on an annotated dataset. First, the cell images need to be indexed by encoding them in a feature space. For this purpose, we propose tailored bio-inspired features relying on the distribution of contrast information. Then, a supervised learning algorithm is proposed for classifying the cells. We carried out experiments on cellular images related to the diagnosis of autoimmune diseases, testing our classification method on the HEp-2 Cells dataset of Foggia et al (CBMS 2010). Results show classification precision larger than 96% on average, thus confirming promising application of our approach to the challenging application of cellular image classification for computer-aided diagnosis.

1. Introduction

Pathologists establish their diagnostics by studying tissue sections, blood samples or punctures. In general, samples are stained with various dyes to visualize cell cytoplasm and nucleus. In addition, immunohistochemistry is used to study specific protein expression. Using these approaches, pathologists observe tissue damage or cell dysfunction like for example, inflammation, neoplasia or necrosis. Abnormal nuclei allow determining cancer grades. Pathologists recognize aberrant shapes of whole cells, organelles, nuclei or staining allowing the classification of the cells. Classical quantification is based on visual counting. Such analysis by one (or several) experimenter is time-consuming and above all poorly reproducible. Furthermore, visual counting is generally performed on a small portion of the sample. A Computer Aided Diagnosis (CAD) system will allow reliable quantification and therefore be a precious tool in diagnostics.

We developed a new classification method for the analysis of the staining morphology of thousands / millions of cells. Our classification process consists in two major steps: first we compute specific bio-inspired descriptors, using contrast information distributions on the already segmented cell. The second step is a boosted supervised nearest neighbors algorithm, UNN [7]. The theory underlying the convergence of this algorithm is a strong advocacy for a careful domain-based tuning of descriptors, a fact clearly borne out from our experiments: our bio-inspired features are sometimes more than 10% more accurate than standard descriptors.

In Section 2 we describe our new Bio-Inspired Feature based on the local contrast information. Section 3 presents UNN and sketches some properties relevant to our paper's topics. Finally, experiments on a dataset of IIF-stained cells are discussed in Section 4.

2. Bio-inspired Contrast-based cell descriptor

The basic idea of bio-inspired descriptor is to get inspiration from the way our visual system operates to analyze the scene. The first transformation undergone by a visual input is performed by the retina.

In fact, ganglion cells, that are the final output of the retina, are first simulated by the local changes of the *illumination*. This information is captured by their receptive fields and transformed to *luminance contrast* intensities. Those receptive fields are like center-surround models. They react to the illumination of either the center or the surround of the ganglion cells and are disabled when illuminating the other one. Such behavior, similar to an edge detector, is modeled by a centered two-dimensional *Difference of Gaussians* (1).

$$DoG_\sigma(x,y) = G_\sigma(x,y) - G_{\alpha\cdot\sigma}(x,y) \qquad (1)$$

Moreover, ganglion cells react to the luminance in different scales, thus adding multi-scale aspect and allowing us to use DoG filters in a scale space.

The basic idea is to compute features inspired from the visual system model and specially from the main characteristics of the retina processing. Such descriptor is well adapted in the case of our cells images since the most discriminative visual feature between categories is the *luminance contrast* in sub-cellular regions. Thus, we define cell descriptors based on the *local contrast* in the cell that we call BIF (Bio-Inspired Features) [1]. The *local contrast* is obtained by a filtering with *Differences of Gaussians* (DoGs) centered at the origin. So that the contrast C_{Im} for each position (x, y) and a given scale s in the image Im is as follows:

$$C_{Im}(x,y,s) = \sum_i \sum_j (Im(i{+}x, j{+}y) \cdot DoG_{\sigma(s)}(i,j)) \qquad (2)$$

We use the DoG described by [2] where the larger Gaussian has three times the standard deviation of the smaller one. After computing these contrast coefficients in (2), we apply a non-linear bounded transfer function, named neuron *firing rates*, used in [10]. This function is written as:

$$R(C) = G \cdot C/(1 + Ref \cdot G \cdot C), \qquad (3)$$

where G is named the contrast gain and Ref is known as the refractory period, a time interval during which a neuron cell *reacts*. The values of those two parameters proposed in [10] to best approximate the retinal system are $G = 2000\,Hz \cdot contrast^{-1}$ and $Ref = 0.005\,s$.

Firing rate coefficients $R(C)$ are encoded on an already segmented cell region. Then, they are quantified into normalized $\mathcal{L}1$ histograms of n-bins for each scale and finally concatenated. Thus our global descriptor's dimension is a multiple of n.

Note that state of the art classical methods such as SIFT descriptors encode gradient directions on square blocks [5].

3. UNN classification

We consider the multi-class problem of automatic cell classification as multiple binary classification problems in the common one-versus-all learning framework [8]. Thus, for each class c, a query image is given a positive (negative) membership with a certain confidence (classification score). Then the label with the maximum score is assigned to the query.

We suppose given a set $\mathcal{S}$ of m annotated images. Each image is a training *example* $(\boldsymbol{o}_i, \boldsymbol{y}_i)$, where $\boldsymbol{o}_i$ is the image feature vector and $\boldsymbol{y}_i \in \{-1, 1\}^C$ denotes the *class vector* that specifies the category membership of the image ($i = 1, 2, ..., m$). In particular, the sign of component y_c gives the positive/negative membership of the example to class c ($c = 1, 2, ..., C$).

In this paper, we propose to use the following leveraged classification rule [7]:

$$h_c^\ell(\boldsymbol{o}) = \sum_{j \sim_k \boldsymbol{o}} \alpha_{jc} y_{jc} \ , \qquad (4)$$

where $\boldsymbol{o}$ denotes the query image, "$j \sim_k \boldsymbol{o}$" denotes the predicate that training example $(\boldsymbol{o}_i, \boldsymbol{y}_i)$ belongs to the k nearest neighbors of $\boldsymbol{o}$, and α_{jc} is a *leveraging coefficient*. We end up with a *weighted* NN voting rule. In the standard k-NN rule, we would have $\alpha_{jc} = 1$.

In order to fit our leveraged classification rule (4) onto training set $\mathcal{S}$, we follow a popular trend in classification which consists in iteratively minimizing a so-called surrogate risk [6]. This risk sums over examples a *loss* with convenient convexity and differentiability properties and which, modulo appropriate scaling, upperbounds the popular zero-one loss. Thus, the surrogate risk brings an upperbound of the empirical risk — hence the "surrogate" name —, and its minimization may be viewed as an approximate primer for the minimization of the empirical risk. This surrogate risk is defined as follows:

$$\varepsilon^\psi\left(h_c^\ell, \mathcal{S}\right) \doteq \frac{1}{m} \sum_{i=1}^{m} \psi\left\{\rho(h_c^\ell, i)\right\} \ , \qquad (5)$$

where ψ, strictly convex and differentiable, defines the loss, and $\rho(h_c^\ell, i) = y_{ic} h_c^\ell(\boldsymbol{o}_i)$ is the multiclass *edge* of classifier h_c^ℓ on training example $\boldsymbol{o}_i$. This edge measures the "goodness of fit" of the classifier on example $(\boldsymbol{o}_i, \boldsymbol{y}_i)$ for class c: it is positive iff the prediction

loss function	δ_j in (6)	w_i in (7)
$\psi^{\exp} \doteq \exp(-x)$	$\frac{1}{2}\log\left(\frac{w_j^{(c)+}}{w_j^{(c)-}}\right)$	$w_i \exp\left(-\delta_j \mathrm{r}_{ij}^{(c)}\right)$
$\psi^{\log} \doteq \log(1+\exp(-x))$	$\log\left(\frac{w_j^{(c)+}}{w_j^{(c)-}}\right)$	$\frac{w_i \exp\left(-\delta_j \mathrm{r}_{ij}^{(c)}\right)}{1-w_i\left(1+\exp\left(-\delta_j \mathrm{r}_{ij}^{(c)}\right)\right)}$
$\psi^{\mathrm{mat}} \doteq -x+\sqrt{1+x^2}$	$\frac{w_j^{(c)+}-w_j^{(c)-}}{2\sqrt{w_j^{(c)+}w_j^{(c)-}}}$	$1-\frac{1-w_i+\sqrt{w_i(2-w_i)}\delta_{jc}r_{ij}^{(c)}}{\sqrt{1+\delta_{jc}^2 w_i(2-w_i)+2(1-w_i)\sqrt{w_i(2-w_i)}\delta_{jc}r_{ij}^{(c)}}}$

Table 1. Three common loss functions and the corresponding expressions for δ_j (6) and w_i (7).

Algorithm 1: Algorithm UNIVERSAL NEAREST NEIGHBORS UNN($\mathcal{S}, \psi$)

Input: $\mathcal{S} = \{(\boldsymbol{o}_i, \boldsymbol{y}_i), i = 1, 2, ..., m\}$, loss ψ;

Let $r_{ij}^{(c)} \doteq \begin{cases} y_{ic}y_{jc} & \text{if} \quad j \sim_k i \\ 0 & \text{otherwise} \end{cases}, \forall i, j, c$;

for $c = 1, 2, ..., C$ **do**

Let $\alpha_{jc} \leftarrow 0, \quad \forall j$;

Let $w_i \leftarrow -\nabla_\psi(0) \in \mathrm{R}_{+*}^m, \quad \forall i$;

for $t = 1, 2, ..., T$ **do**

[I.0] Let $j \leftarrow \text{WIC}(\{1, 2, ..., m\}, t)$;

[I.1] Let $\delta_j \in \mathrm{R}$ solution of:

$$\sum_{i=1}^{m} r_{ij}^{(c)} \nabla_\psi \left(\delta_j r_{ij}^{(c)} + \nabla_\psi^{-1}(-w_i)\right) = 0 \ ; \tag{6}$$

[I.2] $\forall i \ : \ j \sim_k i$, let

$$w_i \leftarrow -\nabla_\psi \left(\delta_j \mathrm{r}_{ij}^{(c)} + \nabla_\psi^{-1}(-w_i)\right) \ ; \tag{7}$$

[I.3] Let $\alpha_{jc} \leftarrow \alpha_{jc} + \delta_j$;

Output: $h_c(\boldsymbol{o}) = \sum_{i \sim_k \boldsymbol{o}} \alpha_{ic} y_{ic}, \quad \forall c \ ;$

agrees with the example's annotation, and its absolute value quantifies a confidence in classification. Some examples of this function that have been implemented and tested in this work, are given in Tab.1. The general version of UNN is shown in Algorithm 1. It is important to notice that there is no closed form for the general solution of (6). Tab.1 provide approximations to these solutions that match the solution when all weights w_i are identical in (6). To keep expressions as compact as possible, we have used the following notations: $w_j^{(c)+} = \sum_{i:\mathrm{r}_{ij}^{(c)}>0} w_i$, $w_j^{(c)-} = \sum_{i:\mathrm{r}_{ij}^{(c)}<0} w_i$, where w_i is a weighting factor depending on the past leveragings. More details on UNN may be found in [7]. A theoretical convergence rate of UNN sheds light on the importance of a careful, domain based crafting of descriptors [7]. Assume that index j returned by WIC satisfies $|w_j^{(c)+}/(w_j^{(c)+} + w_j^{(c)-}) - \frac{1}{2}| \geq \gamma > 0$, which means that example j achieves a classification on class c different from pure random from at least some (eventually very small) constant γ. In this case, after t rounds of leveragings for class c in UNN, the surrogate risk, and hence the empirical risk, of the leveraged k-NN meets:

$$\varepsilon^\psi \left(h_c^\ell, \mathcal{S}\right) \leq \exp(-u\gamma t) \ , \tag{8}$$

for some $u > 0$ on which the choice of descriptors choice play little role [7]. The choice of descriptors however heavily impact γ, as a careful choice may significantly increase the separability between classes, and hence increase γ as well. Between two choices of descriptors, one meeting the γ bound, and a better one ensuring $\gamma' = (1+\varepsilon)\gamma$, one sees that the number of iterations to reach a specific bound in (8) is reduced by a *factor* $(1+\varepsilon)$, and thus may significantly impact the leveraged k-NN. In our experiments, we use a "soft" version of the UNN classifier which considers a logistic estimator for a Bernoulli prior that vanishes with the rank of the neighbors, thus decreasing the importance of farther neighbors: $\hat{p}(j) = \beta_j = (1 + \exp(\lambda(j-1)))^{-1}$. This amounts to redefining (4) as follows:

$$h_c^\ell(\boldsymbol{x_i}) = \sum_{j=1, j \in kNN(i)}^{k} \alpha_{jc}\beta_j y_{jc} \ . \tag{9}$$

4. Experiments

We evaluated our classification approach on the HEp-2 Cells dataset [3] provided by University of Salerno and Campus Bio-Medico of Roma[1]. This database contains 721 images divided into *six* categories as shown in Fig. 1. Cells are already segmented (manually) and both hole images and their corresponding masks are provided in the dataset.

[1]Data available at: `http://mivia.unisa.it/hep2contest/index.shtml`

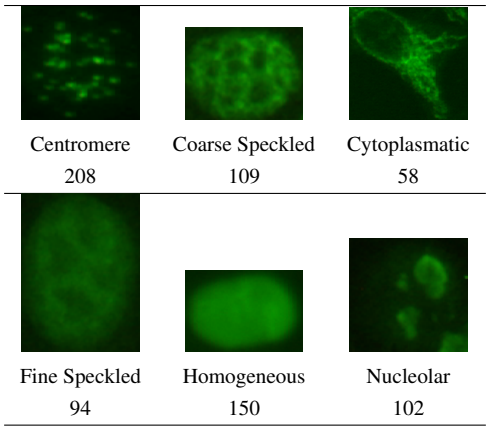

Figure 1. Sample images and the number of elements for each category in the dataset.

In a first step, we extract *Bio-Inspired* features for each *manually* segmented cell according to the cell mask. This version of our feature will be denoted as BIF^s. In a second experiment, we extracted BIF on the whole image of the cell (without segmentation) to test the robustness of those features. We will refer to this version by BIF^a. To better adjust some parameters, such as the dimension, we performed a cross validation system on the number of scales and the number of quantification bins, and we choose using 4 scales with a number of bins equal to 256. Our global features are the concatenation of histograms of 256-bins for each scale. The final dimension of descriptors is then equal to 4×256. We compare our approach to the state of the art SIFT descriptor. We use classical Bag-of-Features [9], with the same dimension 1024, on the dense SIFT provided by [11] which encode gradient directions on a grid of small square blocks of the cellular image.

	UNN^{exp}	UNN^{log}	UNN^{mat}
TP rate	96.16	95.46	94.72
AUC	96.32	95.78	95.25

Table 2. Classification results using the BIF^s descriptor for the three proposed versions of UNN. The first row indicates the True Positive rate (TP rate), and the second one is about the Area Under the roc Curve (AUC).

For the classification task we performed cross validations on 10 random folds. Each fold corresponds to a random split of the dataset such that we train on 50% of the images, while testing on the remaining ones. We evaluated the different versions of UNN in Tab.2.

We compared performances of UNN^{exp} with those of standard SVM, using both BIF and SIFT Bags-of-features (see Tab.3). The reported results of UNN refer to setting $k = 10$ for both training and testing. This value refers to the best performances according to a cross validation on the training set. The same experiment was performed to choose the parameters for the *gaussian* SVM. Note that for the k-NN search we used a fast and efficient software provided by [4]. For BIF descriptor we report experiments on both BIF^s and BIF^a versions of our features. Although BIF^a version performs similar results to BIF^s version, the comparison with SIFT Bags-of-features becomes fair enough to conclude that Bio-Inspired Features are more adapted to such images. In fact, results on Tab.3 display the high discriminative ability of the proposed Bio-Inspired Feature, which allows for classification precision generally larger than 90%, up to almost 100% (on the "Coarse Speckled" and "Cytoplasmatic" classes). In addition, the precision obtained with such specific descriptor outperforms the standard SIFT bag-of-features by at least 14% in terms of True Positive rate (TP rate) and 9% in terms of Area Under the roc Curve (AUC). Furthermore, the most interesting results are those obtained using BIF^a, since in real cases an automatic segmentation process on cellular images is poorly reproducible. Those results (columns in bold in Tab.3) shows not only the efficiency of the feature but also the precision of our UNN algorithm which remains relevant (in terms of TP rate and AUC), comparable and even better than state-of-the-art SVM. For instance, notice the improvement of UNN over SVM on the "Coarse Speckled" class (4.5% of gap), while SVM is the best performing method on the "Homogeneous" class (3% of gap). Besides comparing very favorably

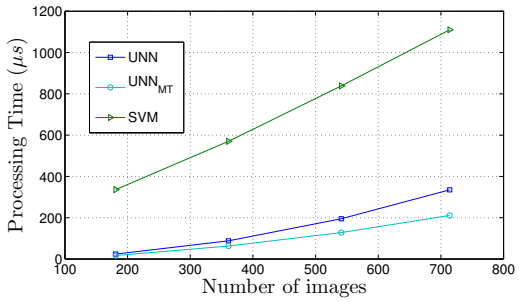

Figure 2. Processing time of the training step for both UNN, SVM and multi-thread version of UNN.

TP rate

	UNN			SVM		
	BIF^a	BIF^s	SIFT	BIF^a	BIF^s	SIFT
Centro.	**96.05**	96.15	85.00	**97.01**	97.40	88.07
C.Spec.	**99.62**	97.59	69.81	**95.00**	97.03	71.29
Cytop.	**100.0**	100.0	99.65	**100.0**	100.0	97.93
F.Spec.	**93.82**	95.95	61.27	**94.25**	94.46	58.93
Homog.	**90.26**	91.20	91.86	**93.46**	94.00	88.93
Nucleo.	**97.45**	96.07	87.25	**97.64**	97.45	88.03
avg TPr	**96.20**	96.16	82.47	**96.23**	96.72	82.20

AUC

	UNN			SVM			
	BIF^a	BIF^s	SIFT	BIF^a	BIF^s	SIFT	
Centro.	**95.48**	95.68	92.63	**97.86**	96.62	92.03	Centro.
C.Spec.	**98.54**	97.24	86.70	**94.23**	95.40	79.00	C.Spec.
Cytop.	**99.64**	99.73	97.82	**99.39**	99.02	93.15	Cytop.
F.Spec.	**93.54**	95.61	63.35	**89.56**	91.82	59.26	F.Spec.
Homog.	**93.42**	94.79	91.06	**97.04**	97.78	91.39	Homog.
Nucleo.	**97.74**	94.89	92.35	**94.94**	98.66	92.59	Nucleo.
avg AUC	**96.39**	96.32	87.32	**95.50**	96.55	84.57	avg AUC

Table 3. Evaluations of UNN and SVM using both BIF^a(on whole images), BIF^s (on *manually* segmented cells) and SIFT Bag-of-features. The table on the left gives the TP rate per class. The last row shows the average TP rate. The same evaluation for the table on the right but using the Area Under the roc Curve (AUC). The best performance for each category is given in blue and the second one in green.

with state-of-the-art approaches, our UNN method enables much faster classification. Fig. 2, shows typical processing time for UNN and SVM and UNN achieves speedups of roughly 3 to 5 over SVM. UNN benefits from straightforward multi-thread implementation (UNN$_{MT}$) in addition to the fast k-NN search algorithm. This makes the processing furthermore faster. Therefore our Bio-Inspired UNN algorithm provides the best Precision/Time trade-off.

5. Conclusion

In this paper, we have presented a novel algorithm for automatic supervised classification of cellular images. First of all, our method relies on extracting highly discriminative BIF descriptors based on the distribution of contrast information on segmented cells. Then, we use a boosting algorithm, called UNN, for learning the most relevant prototypical samples that are to be used for predicting the class of unlabeled cellular images according to a leveraged k-NN rule. We evaluated BIF and UNN performances on the HEp-2 Cells dataset (manually segmented and annotated). Although being the early results of our methodology for such a challenging application, performances are really satisfactory (average global precision of 96%) and display the importance on UNN of careful domain-based tunings of descriptors, a fact suggested by theory but so far never experimentally tackled.

References

[1] W. Bel haj ali, P. Piro, L. Crescence, D. Giampaglia, O. Ferhat, J. Darcourt, T. Pourcher, and M. Barlaud. A bio-inspired learning and classification method for subcellular localization of a plasma membrane protein. In *International Conference on Computer Vision Theory and Applications (VISAPP 2012)*, 2012.

[2] D. J. Field. What is the goal of sensory coding? *Neural Computation*, 6(4):559–601, 1994.

[3] P. Foggia, G. Percannella, P. Soda, and M. Vento. Early experiences in mitotic cells recognition on hep-2 slides. In *CBMS'10*, pages 38–43, 2010.

[4] H. Jégou, M. Douze, and C. Schmid. Product quantization for nearest neighbor search. *IEEE Trans. PAMI*, 33(1):117–128, jan 2011. to appear.

[5] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[6] R. Nock and F. Nielsen. Bregman divergences and surrogates for learning. *IEEE Trans. PAMI*, 31(11):2048–2059, 2009.

[7] P. Piro, R. Nock, F. Nielsen, and M. Barlaud. Leveraging k-nn for generic classification boosting. *Neurocomputing*, 80:3–9, March 2012.

[8] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37:297–336, 1999.

[9] J. Sivic and A. Zisserman. Video google: Efficient visual search of videos. In J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors, *Toward Category-Level Object Recognition*, volume 4170 of *Lecture Notes in Computer Science*, pages 127–144. Springer Berlin / Heidelberg, 2006. 10.1007/11957959_7.

[10] R. Van Rullen and S. J. Thorpe. Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput*, 13(6):1255–1283, June 2001.

[11] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. `http://www.vlfeat.org/`, 2008.