Vertex Proportion Loss for Multi-Class Cell Detection from Label Proportions

Carolina Pacheco (✉)[1], Florence Yellin[2], René Vidal[3], and Benjamin Haeffele[1]

[1] Mathematical Institute for Data Science (MINDS),
Johns Hopkins University, Baltimore, MD, USA. [cpachec2, bhaeffele]@jhu.edu
[2] Kitware Inc., Clifton Park, NY, USA.
[3] Center for Innovation in Data Engineering and Science (IDEAS), University of Pennsylvania, Philadelphia, Pennsylvania, USA.

Abstract. Learning from label proportions (LLP) is a weakly supervised classification task in which training instances are grouped into bags annotated only with class proportions. While this task emerges naturally in many applications, its performance is often evaluated on bags generated artificially by sampling uniformly from balanced, annotated datasets. In contrast, we study the LLP task in multi-class blood cell detection, where each image can be seen as a "bag" of cells and class proportions can be obtained using a hematocytometer. This application introduces several challenges that are not appropriately captured by the usual LLP evaluation regime, including variable bag size, noisy proportion annotations, and inherent class imbalance. In this paper, we propose the Vertex Proportion loss, a new, principled loss for LLP, which uses optimal transport to infer instance labels from label proportions, and a Deep Sparse Detector that leverages the sparsity of the images to localize and learn a useful representation of the cells in a self-supervised way. We demonstrate the advantages of the proposed method over existing approaches when evaluated in real and synthetic white blood cell datasets.

1 Introduction

Large, annotated datasets played a critical role in the early success of deep models. Since then, extending this success to unsupervised and weakly supervised regimes has been an active focus of research. One example is *Learning from Label Proportions (LLP)*, a weakly supervised task that aims to learn an instance classifier without instance-level annotations. In particular, LLP assumes that the classifier has access to bags of instances during training, each one annotated only with the proportion of instances corresponding to each class (e.g., 70% class A, 30% class B). This LLP paradigm not only reduces the annotation burden but also arises naturally in diverse applications. For example, in vision-based blood count, the goal is to detect and classify all the blood cells (instances) in an image (bag). Annotating the location and class of each cell becomes prohibitively expensive given the large number of cells and required expertise; however, class proportion information can be obtained using a hematocytometer [31]. Similarly,

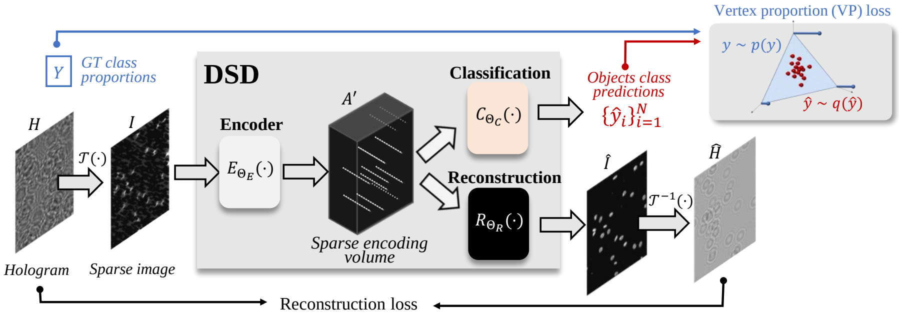


Fig. 1. DSD performs multi-class cell detection by learning a structured latent representation of the image, suitable for object classification and image reconstruction. The classification task is weakly supervised by label proportions via the proposed VP loss, while the reconstruction task is self-supervised. For lensless imaging, a physics-based, sparsifying transformation $\mathcal{T}(\cdot)$ is incorporated to the model.

when analyzing lung images of emphysema patients, one might be interested in predicting segmentation masks (instances) corresponding to damaged tissue, but due to annotation cost only have access to the proportion of diseased tissue in each image (bag) [4]. Other relevant LLP applications include assisted reproductive technology [12,11] and vision-based sperm cell classification [7].

While several approaches to address the LLP task have been explored in the past, existing methods rely on ad-hoc heuristics, are very slow to compute, require precise hyperparameter tuning, or have numerous trivial solutions. Recent methods perform extensions to multi-class, high-dimensional tasks like image classification [17,24,30,16,26,6,18,15]; however, their performance in real-world LLP settings remains unknown since they have been consistently evaluated in an artificial LLP scenario. Namely, images are uniformly sampled from balanced classification datasets (MNIST, CIFAR-10, CIFAR-100, among others) to create bags of a fixed size, and their class proportions are obtained by aggregating the ground-truth (GT) instance labels. This evaluation setting fails to replicate key aspects observed in real-world applications such as class imbalance, noisy label proportions, and variable bag size.

We aim to advance the understanding of LLP by studying it in the context of a relevant application: multi-class cell detection in lensless imaging. We focus on white blood cell (WBC) subtype classification, which is critical in the assessment of infections and the status of the immune system in general [2]. Among other challenges, this task presents a strong class imbalance, with granulocytes and monocytes representing more than 50% and less than 12% of the WBCs, respectively [5]. Moreover, WBC proportions from complete blood count (CBC) reports are imprecise for this task, since they are not obtained from the same field of view as our images, and the detection process incorporates additional noise. In this context, the contributions of this paper are the following:

1. We propose a principled LLP loss, the *vertex proportion (VP) loss*, which is based on optimal transport (OT) and propagates global annotations to instance labels as latent variables. Unlike OT-based pseudo-labeling approaches [6,18,15], the VP loss computation is simple as it does not require additional hyperparameters, alternating updates, or postprocessing steps.
2. We introduce the *deep sparse detector (DSD)*, a self-supervised model that localizes cells and learns useful representations to classify them in a unified framework. Unlike the greedy state-of-the-art detector for this problem [31], DSD can detect hundreds of cells in an image with just a single forward pass, and it does not rely on alternative data sources for training.
3. We study, for the first time, the *WBC subtype classification task as an LLP problem*. We evaluate our method along with other common LLP approaches in a real dataset, and also generate a new, synthetic dataset [4] which allows for the computation of detailed metrics. Unlike artificial scenarios [6], in this case we observe that inferring instance-level labels during training is critical.

2 Proposed approach

We address this weakly supervised multi-class cell detection task through a unified architecture, the DSD model in Fig. 1, and divide the learning problem into self-supervised localization and LLP classification. Sec. 2.1 introduces the LLP problem and our VP loss to learn from proportions (given a bag of cell embeddings), and Sec. 2.2 explains how DSD predicts cell locations and embeddings.

2.1 Learning from label proportions (LLP)

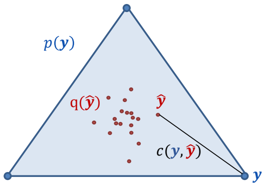
Let $\left\{(\mathbf{x}_i, \mathbf{y}_i)\right\}_{i=1}^{N}$ be the set of N cells present in an image (bag), where $\mathbf{x}_i \in \mathbb{C}^d$ is the embedding associated with the i-th object (we use complex-valued features because they arise naturally in lensless imaging), and $\mathbf{y}_i \in \mathbb{R}^K$ is a one-hot vector indicating its GT class (out of K classes). The classification task aims to find a function $\mathcal{C}_{\Theta_C}$ that correctly estimates the class labels, i.e., $\mathcal{C}_{\Theta_C}(\mathbf{x}_i) := \hat{\mathbf{y}}_i \approx \mathbf{y}_i, \forall i$. This is usually achieved by minimizing the cross-entropy loss; however, its computation requires cell-level class annotations $\mathbf{y}_i$, which are not available in LLP. In contrast, we need to learn from class proportions $Y = \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i$.

As an initial approach to this problem [1,6], one might consider the Kullback-Leibler divergence (KL-div) loss between the predicted proportions and the true proportions, given as $D_{KL}\left(Y \parallel \hat{Y}\right) = \sum_{k=1}^{K} Y^k \cdot \log\left(\frac{Y^k}{\hat{Y}^k}\right)$, with $\hat{Y} = \frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{y}}_i$. However, note that the $D_{KL}(Y||\hat{Y})$ loss is somewhat ill-posed for the LLP problem as it is not only minimized in the desired scenario ($\hat{\mathbf{y}}_i = \mathbf{y}_i, \forall i$), but also in cases where no information is learned about the class of a particular instance (e.g., $\hat{\mathbf{y}}_i = Y, \forall i$ is a global optimum). This is a common disadvantage of losses

[4] This dataset and code relevant to this paper can be found in GitHub

that operate directly on the aggregated information $\hat{Y}$ (i.e., bag-level losses). In the literature this has not been reported as a problem [6,1,17], likely because they have studied artificial LLP scenarios where bags of instances can be randomly drawn from large datasets, with direct control over the bag sizes, thus providing enough bag diversity for training. For the application of LLP to real-world settings, we propose a new loss, the **VP loss**, designed to infer instance-level labels during training by matching the distributions of predicted and GT class labels, as opposed to relying only on aggregated information.

VP loss to learn from proportions. We are interested in matching the distribution of predicted labels $q(\hat{\mathbf{y}})$ to the distribution of GT labels $p(\mathbf{y})$. In supervised learning, annotations directly define an assignment between the predicted label $\hat{\mathbf{y}}_i$ and the GT label $\mathbf{y}_i$ of each instance, and learning is achieved by minimizing the empirical expectation of a cost $c(\mathbf{y}, \hat{\mathbf{y}})$, that can be, for example, the cross-entropy loss. In LLP we aim to do the same, yet the label assignment is unknown. What we do know, however, is the distribution of $p(\mathbf{y})$. Namely, given GT class proportions Y, we can write $p\left(\mathbf{y}\right) = \sum_{k=1}^{K} Y^k \cdot \delta(\mathbf{y} - \mathbf{e}_k)$, where $\mathbf{e}_k \in \mathbb{R}^K$ corresponds to a canonical vector that is non-zero in its k-th entry, and $\delta(\cdot)$ represents the delta function that is non-zero at the origin. Since we assume that each cell belongs to one class, the support of $p(\mathbf{y})$ is concentrated on the vertices of the simplex (blue dots in Fig. 2) and the corresponding probability mass associated with each vertex (class) is defined by the entries of Y.



$p(\boldsymbol{y})$
$q(\hat{\boldsymbol{y}})$
$\hat{\boldsymbol{y}}$
$c(\boldsymbol{y}, \hat{\boldsymbol{y}})$
$\boldsymbol{y}$

Fig. 2. Distributions of interest. $p(\mathbf{y})$: GT labels, $q(\hat{\mathbf{y}})$: class predictions, and $c(\mathbf{y}, \hat{\mathbf{y}})$: cost for moving a unit of mass from $\mathbf{y}$ to $\hat{\mathbf{y}}$.

OT aims to find a transportation plan (i.e., assignment) of minimal cost to match the probability mass between two distributions [25]. Although it is a well-studied problem with numerous applications, it is intractable to solve in the general case. However, here the structure of $p(\mathbf{y})$ significantly simplifies the formulation. In particular, given a transport cost function $c(\cdot, \cdot)$, GT proportions Y, and predicted labels $\{\hat{\mathbf{y}}_i\}_{i=1}^{N}$, after approximating an analytic expectation with an empirical expectation based on samples $\{\hat{\mathbf{y}}_i\}_{i=1}^{N} \sim q$, the OT problem reduces to (see details in the supplement) the following linear program (LP)

$$\max_{\Phi \in \mathbb{R}^K, \mathbf{s} \in \mathbb{R}^N} \langle Y, \Phi \rangle - \frac{1}{N} \langle \mathbf{s}, \mathbf{1}_N \rangle \quad \text{s.t.} \quad s_i \geq \phi_k - c(\mathbf{e}_k, \hat{\mathbf{y}}_i), \forall (i,k) \in [N] \times [K], \tag{1}$$

where $\mathbf{1}_N \in \mathbb{R}^N$ is the vector of all ones, ϕ_k is the k-th entry of Φ, s_i is the i-th entry of $\mathbf{s}$, and $[N]$ is the set of integers from 1 to N. The optimal label assignment for the i-th cell is given by $\mathbf{e}_{k^*}$, with $k^* \in \arg\max_k \{\phi_k^* - c(\mathbf{e}_k, \hat{\mathbf{y}}_i)\}$ and ϕ_k^* the k-th entry of the optimal value of Φ in Eq. (1).

Now, we would like to use gradient-based optimization to train our classifier $\mathcal{C}_{\Theta_C}(\cdot)$, using as a loss function the expectation of $c(\cdot, \cdot)$ given the optimal assignments. In particular, we need to compute the gradient of the loss with respect to the output of the network $\hat{\mathbf{y}}_i$. Note that (i) the value of this loss is by definition the same as the optimal value of OT, i.e., the solution to Eq. (1), and (ii) the learnable parameters in Eq. (1) appear only in the evaluation of the costs $c(\mathbf{e}_k, \hat{\mathbf{y}}_i)$, with $\hat{\mathbf{y}}_i = \mathcal{C}_{\Theta_C}(\mathbf{x}_i)$. Therefore, we need to compute the gradient of Eq. (1) with respect to $c(\mathbf{e}_k, \hat{\mathbf{y}}_i)$, and then the gradient of $c(\mathbf{e}_k, \hat{\mathbf{y}}_i)$ with respect to $\hat{\mathbf{y}}_i$ can be obtained in closed form or via auto-differentiation. Given the linear structure of the problem, *the gradient of the loss with respect to the constraints $c(\boldsymbol{e}_k, \hat{\boldsymbol{y}}_i)$ conveniently equals the optimal value of the dual variable* [20], and thus it is obtained directly from the computation of the loss, without having to explicitly compute or store the assignments.

Connections to other LLP approaches. Current LLP approaches can be divided into three categories: new losses [1,6]; pseudo-labeling strategies to alternate with supervised losses [6,18,15]; and representation learning approaches [17,24,30,16,26,19]. Our loss belongs to the first group, yet it enjoys advantages similar to pseudo-labeling as it also infers instance-level information for training.

2.2 Deep Sparse Detector (DSD)

The LLP approach described in Sec. 2.1 assumes that cells have already been detected in the image and that the corresponding features $\mathbf{x}_i$ have been extracted. Here, we focus on the problem of detecting cells and extracting features. To do so, DSD leverages the sparsity of the imaged specimen, i.e. blood, to generate a structured latent representation of the image, referred to as a *sparse encoding volume*. This representation is spatially sparse, and its local features are discriminative for cell classification. In absence of detailed annotations, we incorporate reconstruction as an auxiliary task, which has been found useful for unsupervised detection [13,31,21]. While this design is tailored to sparse images, which can be reconstructed from localized feature maps, lensless imaging captures the intensity of the diffraction pattern of the objects as *holograms*, and thus even if the specimen is sparse, the recorded hologram is not. Fortunately, there exist physics-based models for the diffraction process [14], which can be well approximated by a linear transformation $I = \mathcal{T}(H)$, where H is a real-valued hologram and I is a complex-valued image (see Fig. 3). Therefore, we incorporate $\mathcal{T}$ as the first layer of DSD, and its inverse as the last layer of the reconstruction head. This strategy can be applied to other cases where images are approximately sparse under invertible, differentiable transformations. Due to the limited resolution of lensless images and the small cell size, this detection task is closer to keypoint estimation than to object detection, as the spatial localization of a cell can be characterized by the location of its center. Lacking supervision, we enforce sparsity by introducing structure into the model as described next.

Deep Sparse Encoder. Given an image $I \in \mathbb{C}^{L\times L}$ containing N objects, the goal of the encoder is to generate a sparse volume as $A = E_{\Theta_E}(I)$, where E_{Θ_E} corresponds to a sequence of complex-valued CNN layers followed by ReLU non-linearities, and $A \in \mathbb{C}^{d\times L\times L}$ is the encoding volume composed of d complex-valued feature maps. Unlike other works relying on heatmap annotations for training [29,28,8], our encoder encourages spatially "peaked" representations by means of a highly local version of softmax [3]. More specifically, we compute a heatmap at pixel (x, y) as $H_A(x, y) = \text{LocalSoftMax}(\|\lambda A(\cdot, x, y)\|_2^2)$, where $\|A(\cdot, x, y)\|_2^2$ is the "strength" of the detection at pixel (x, y) defined as the L_2 norm squared of the encoding vector at (x, y), $\lambda \in \mathbb{R}^+$ is a hyperparameter that controls the decay rate from the largest value, and the normalization of the softmax is computed locally in a patch of size $m \times m$ centered at (x, y). The LocalSoftMax equation is easily implemented with a sequence of point-wise operations for the numerator, and then pooling for the denominator. To obtain the sparsest possible representation for each object, we also apply thresholding and non-maximum suppression (NMS) in small patches. We utilize the resulting sparse mask as $A' = M_A \odot A$, where $\odot$ denotes point-wise multiplication, and $M_A(\cdot) \in \mathbb{R}^{d \times L \times L}$ represents a spatial mask which is constant across the first dimension (features), and at each pixel contains the thresholded, NMS version of H_A. From this sparse encoding volume, $A' \in \mathbb{C}^{d \times L \times L}$, one can extract both the location of the detected objects from the spatial support of A', and their encodings as $\mathbf{x}_i = A'(\cdot, x_i, y_i) \in \mathbb{C}^d$, for (x_i, y_i) in the support of A'.

Reconstruction head. We aim to reconstruct the input image from the output of the encoder as $\hat{I} = R_{\Theta_R}(A')$. Inspired by the common practice of generating GT keypoint heatmaps by convolving a sparse mask with small, smoothing filters [23,27], we use a sequence of complex-valued CNN layers with small filters to parameterize R_{Θ_R}, thus limiting the spatial extent of the reconstructed cells. We train the network with the Frobenius norm squared as a loss. Since the recorded images correspond to holograms, we apply the inverse of the physics-based transformation $\hat{H} = \mathcal{T}^{-1}(\hat{I})$ to compute the loss in the hologram domain.

Classification head. From a cell encoding $\mathbf{x}_i$, the classifier aims to correctly predict its class. Assuming that objects from distinct classes look different, DSD encodings are expected to contain somewhat disentangled class information because they are trained to reconstruct the appearance of different cells. We thus use a simple classifier (fully connected layers followed by ReLU non-linearities and a softmax) and train it from label proportions, as described in Sec. 2.1.

3 Experiments

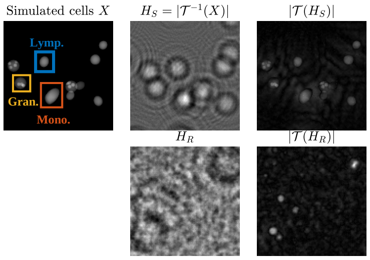


Fig. 3. Simulated (top) versus Real (bottom) image crops. Synthetic cells (left), holograms (middle), and holograms propagated to the object plane via $\mathcal{T}$ (right).

Datasets. We evaluate our method in a real [31] and a new synthetic WBC holographic dataset. Fig. 3 shows examples from both. The real dataset contains images from 33 donors, and it is annotated with the approximated proportions of granulocytes, lymphocytes, and monocytes ($K = 3$) for each donor. In this dataset, the main evaluation metric corresponds to *mean absolute error in the proportion prediction* after hard assignment [24], but we also use WBC concentration as a proxy for detection evaluation. The synthetic dataset was generated modeling the geometric and optical properties of the WBC subtypes. It simulates data from 500 subjects, divided in train, validation, and test sets (200/100/200). The evaluation metrics correspond to precision, recall and f1-score for detection, as well as classification accuracy.

LLP classification results. We implement two variants of the proposed VP loss by using the squared loss (VP-L2) and the cross-entropy loss (VP-CE) as cost functions in the OT problem. We compare them to bag-level losses (MSE and KL-div [6]) and pseudo-labeling approaches (Feature-Label Matching (FLM) [30], Prototypic Clustering (PC) [15], LLP-PLOT [18]) under the same model, optimizer, learning rate, and number of epochs (see supplement for implementation details). We first train the encoder with the reconstruction loss, and then the classifier. As usual, instance-level approaches are initialized by pretraining with the KL-div, and hyperparameters are chosen as recommended by the authors. Tables 1 and 2 report results for synthetic and real data, respectively. *Bag-level losses fail to learn discriminative features for synthetic data*, generating high-entropy predictions (0.841 and 0.790 for MSE and KL-div, resp.) close to the global class proportions (0.816 avg. entropy). In contrast, *VP-L2 and VP-CE lead to predictions with significantly lower entropy* (0.162 and 0.196, respectively). Similar trends hold for real data, where bag-level losses and simple pseudo-labeling methods fail to recover cells from the least populated class.

Our proposed loss, as well as LLP-PLOT [18], consistently outperform the KL-div baseline, with *VP-CE achieving the best instance-level accuracy* for synthetic data, and *VP-L2 the smallest donor-level proportion prediction error* for real data. LLP-PLOT performs similarly to the VP loss for the best selection of parameters, however it is very sensitive to the choice of regularizer weight γ and postprocessing variants. Moreover, the slow convergence of the Sinkhorn algorithm leads to a significant increase in training time (see Fig. 4). In contrast, the *VP-CE loss achieves similar or better performance without additional hyperparameters and reduces the training time by one order of magnitude.* Additionally, experiments confirm the *complementary nature of the VP loss with representation learning LLP methods* [17,30,19] (see details in the supplement).

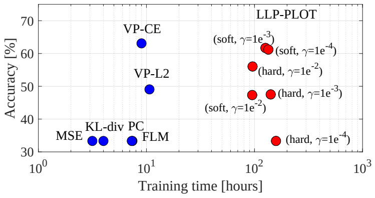


Fig. 4. Instance-level accuracy vs training time in synthetic data. Red circles represent LLP-PLOT [16], for which different variants are explored (regularizer weight γ and postprocessing strategies)

Table 1. Classification accuracy in *synthetic data* for LLP approaches.

	MSE	**KL-div**	**FLM** [30]	**PC** [15]	**LLP-PLOT** [18]	**VP-L2**	**VP-CE**
Lymp.	0.00	0.00	0.00	0.00	65.41	59.35	66.80
Mono.	0.00	0.00	0.00	0.00	37.20	6.45	39.23
Gran.	100.00	100.00	100.00	100.00	82.48	84.80	83.24
Average	33.33	33.33	33.33	33.33	61.70	50.20	**63.09**

Table 2. Mean absolute error of proportion prediction in *real data* for LLP approaches.

	MSE	**KL-div**	**FLM** [30]	**PC** [15]	**LLP-PLOT** [18]	**VP-L2**	**VP-CE**
Lymp.	6.94	6.21	28.91	40.67	5.84	5.37	4.71
Mono.	8.07	7.96	8.94	9.09	5.09	5.09	5.43
Gran.	6.69	9.51	37.85	43.87	6.37	6.26	7.11
Average	7.23	7.89	25.23	31.21	5.77	**5.58**	5.75

We also compare our approach to CSC priors [31], the state of the art for multi-class cell detection in the real WBC dataset. It achieves 5.97 mean absolute error, which is similar to the performance of VP-L2 and VP-CE in Table 2. However, CSC priors exhibits several practical drawbacks: it requires images of purified data obtained through specialized biochemical processes; it relies on a large CBC database of 300K patients to model proportion priors; it requires reconstruction as preprocessing; and it does not learn a representation of the cells suitable for downstream tasks. In synthetic data, we outperform CSC priors for proportion prediction with 4.90 and 4.25 absolute error for VP-L2 and VP-CE, respectively compared to 8.95 for CSC priors.

Comparison to proportion prediction methods. We compare the proposed detection-based approach to regression and classification approaches for class proportion prediction (given oracle detections) in real data. We evaluate both, architectures used in the literature for these tasks, as well as our encoder. Implementation details can be found in the supplement, while Table 3 summarizes their performance. First, regression approaches have limited performance, probably due to the small amount of annotations, as they tend to predict the average proportions of the training data. Thus, the *introduction of structure in the form of detection is beneficial.* Second, our encoder outperforms the ones proposed in the literature, even when both are randomly initialized and trained with the same losses. Therefore, *considering data availability in the architectural design is advantageous.* Third, initializing our encoder with weights learned by the reconstruction task boosts performance (compare pretrained (P) to randomly initialized (R) cases in Table 3). Thus, *there is a positive contribution of using reconstruction as an auxiliary task.*

Table 3. Mean absolute error of proportion prediction in *real data.* (Init: initialization, R: random, P: pretrained with rec. loss). Classifier and encoder are jointly optimized.

Type	**Encoder**	**Init.**	**Loss**	**Error**
Direct regression	ResNet-152 [29]	R	KL-div	8.5
	Ours	R	KL-div	7.6
	Ours	P	KL-div	**7.1**
Heatmap regression	C-FCRN [10]	R	KL-div	8.3
	Ours	R	KL-div	7.9
	Ours	P	KL-div	**7.0**
Classification (oracle detections)	Le Net [9]	R	KL-div	11.5
	Le Net [9]	R	VP-L2	8.4
	Le Net [9]	R	VP-CE	8.2
	Ours	R	KL-div	9.7
	Ours	R	VP-L2	7.3
	Ours	R	VP-CE	7.6
	Ours	P	KL-div	9.3
	Ours	P	VP-L2	6.3
	Ours	P	VP-CE	**6.1**

Detection results. We compare the detection performance of the proposed DSD to *Cellpose* [22] (a popular pretrained model for general cell segmentation); (2) a *baseline* approach (which applies local softmax and NMS directly on the absolute value of the input image); and (3) *CSC priors* [31], previously proposed for cell detection in lensless imaging. Detailed results can be found in the supplement, with the following high-level findings: (i) Cellpose obtains the lowest performance across all metrics, which supports the need of specialized models for lensless imaging, (2) our model outperforms the baseline and achieves the best precision, (3) CSC priors reaches the best performance in most metrics, yet at the expense of additional data labeling and computational complexity.

4 Conclusion

We study LLP in a real-world application in which instances are naturally grouped into bags and (noisy) proportion annotations are easily obtained. Results show that inference of instance-level information is critical during training. We propose a self-supervised detector and an OT-based loss that achieves state-of-the-art results for weakly supervised classification and class proportion prediction, with practical advantages with respect to prior work. **Limitations.** Our detector trades recall for computational efficiency, so it might miss relevant objects. Also, it relies on the assumption that images are sparse under a known, differentiable transformation, thus it is not suitable for dense images in general.

Acknowledgments. This work is supported by NIH NIA #1R01AG067396 and NIH NIAID #1R21AI169363. CP was supported by Amazon under the JHU-Amazon Initiative for Interactive AI (AI2AI).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ardehaly, E.M., Culotta, A.: Co-training for demographic classification using deep learning from label proportions. In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW). pp. 1017–1024. IEEE (2017)
2. Barnes, P., McFadden, S., Machin, S., Simson, E., et al.: The international consensus group for hematology review: suggested criteria for action following automated cbc and wbc differential analysis. Laboratory hematology: official publication of the International Society for Laboratory Hematology **11**(2), 83–90 (2005)
3. Barroso-Laguna, A., Riba, E., Ponsa, D., Mikolajczyk, K.: Key. net: Keypoint detection by handcrafted and learned cnn filters. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5836–5844 (2019)
4. Bortsova, G., Dubost, F., Ørting, S., Katramados, I., Hogeweg, L., Thomsen, L., Wille, M., de Bruijne, M.: Deep learning from label proportions for emphysema quantification. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11. pp. 768–776. Springer (2018)
5. Ciesla, B.: Hematology in practice. Fa Davis (2018)
6. Dulac-Arnold, G., Zeghidour, N., Cuturi, M., Beyer, L., Vert, J.P.: Deep multi-class learning from label proportions. arXiv preprint arXiv:1905.12909 (2019)

7. González-Castro, V., Alaiz-Rodríguez, R., Fernández-Robles, L., Guzmán-Martínez, R., Alegre, E.: Estimating class proportions in boar semen analysis using the hellinger distance. In: Trends in Applied Intelligent Systems: 23rd International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2010, Cordoba, Spain, June 1-4, 2010, Proceedings, Part I 23. pp. 284–293. Springer (2010)
8. Guo, Y., Stein, J., Wu, G., Krishnamurthy, A.: Sau-net: A universal deep network for cell counting. In: Proceedings of the 10th ACM international conference on bioinformatics, computational biology and health informatics. pp. 299–306 (2019)
9. Habibzadeh, M., Krzyżak, A., Fevens, T.: White blood cell differential counts using convolutional neural networks for low resolution images. In: Artificial Intelligence and Soft Computing: 12th International Conference, ICAISC 2013, Zakopane, Poland, June 9-13, 2013, Proceedings, Part II 12. pp. 263–274. Springer (2013)
10. He, S., Minn, K.T., Solnica-Krezel, L., Anastasio, M.A., Li, H.: Deeply-supervised density regression for automatic cell counting in microscopy images. Medical Image Analysis **68**, 101892 (2021)
11. Hernández-González, J., Inza, I., Crisol-Ortíz, L., Guembe, M.A., Iñarra, M.J., Lozano, J.A.: Fitting the data from embryo implantation prediction: Learning from label proportions. Statistical methods in medical research **27**(4), 1056–1066 (2018)
12. Hernández-González, J., Inza, I., Lozano, J.A.: Learning bayesian network classifiers from label proportions. Pattern Recognition **46**(12), 3425–3440 (2013)
13. Hou, L., Nguyen, V., Kanevsky, A.B., Samaras, D., Kurc, T.M., Zhao, T., Gupta, R.R., Gao, Y., Chen, W., Foran, D., et al.: Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images. Pattern recognition **86**, 188–200 (2019)
14. Kim, M.K.: Principles and techniques of digital holographic microscopy. SPIE reviews **1**(1), 018005 (2010)
15. La Rosa, L.E.C., Oliveira, D.A.B.: Learning from label proportions with prototypical contrastive clustering. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36 (2), pp. 2153–2161 (2022)
16. Liu, J., Qi, Z., Wang, B., Tian, Y., Shi, Y.: Self-llp: Self-supervised learning from label proportions with self-ensemble. Pattern Recognition **129**, 108767 (2022)
17. Liu, J., Wang, B., Qi, Z., Tian, Y., Shi, Y.: Learning from label proportions with generative adversarial networks. Advances in neural information processing systems **32** (2019)
18. Liu, J., Wang, B., Shen, X., Qi, Z., Tian, Y.: Two-stage training for learning from label proportions. In: Zhou, Z.H. (ed.) Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21. pp. 2737–2743. International Joint Conferences on Artificial Intelligence Organization (8 2021). https://doi.org/10.24963/ijcai.2021/377, https://doi.org/10.24963/ijcai.2021/377, main Track
19. Nandy, J., Saket, R., Jain, P., Chauhan, J., Ravindran, B., Raghuveer, A.: Domain-agnostic contrastive representations for learning from label proportions. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. pp. 1542–1551 (2022)
20. Rockafellar, R.T.: Conjugate duality and optimization. Society for Industrial and Applied Mathematics (1974)
21. Sinha, A., Lee, J., Li, S., Barbastathis, G.: Lensless computational imaging through deep learning. Optica **4**(9), 1117–1125 (Sep 2017). https://doi.org/10.1364/OPTICA.4.001117, https://opg.optica.org/optica/abstract.cfm?URI=optica-4-9-1117

22. Stringer, C., Wang, T., Michaelos, M., Pachitariu, M.: Cellpose: a generalist algorithm for cellular segmentation. Nature methods **18**(1), 100–106 (2021)
23. Tompson, J., Goroshin, R., Jain, A., LeCun, Y., Bregler, C.: Efficient object localization using convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 648–656 (2015)
24. Tsai, K.H., Lin, H.T.: Learning from label proportions with consistency regularization. In: Asian Conference on Machine Learning. pp. 513–528. PMLR (2020)
25. Villani, C.: Topics in optimal transportation, vol. 58. American Mathematical Soc. (2021)
26. Wang, B., Sun, Y., Tong, Q.: Llp-aae: Learning from label proportions with adversarial autoencoder. Neurocomputing **537**, 282–295 (2023)
27. Wei, S.E., Ramakrishna, V., Kanade, T., Sheikh, Y.: Convolutional pose machines. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 4724–4732 (2016)
28. Xie, W., Noble, J.A., Zisserman, A.: Microscopy cell counting and detection with fully convolutional regression networks. Computer methods in biomechanics and biomedical engineering: Imaging & Visualization **6**(3), 283–292 (2018)
29. Xue, Y., Ray, N., Hugh, J., Bigras, G.: Cell counting by regression using convolutional neural network. In: Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part I 14. pp. 274–290. Springer (2016)
30. Yang, H., Zhang, W., Lam, W.: A two-stage training framework with feature-label matching mechanism for learning from label proportions. In: Asian Conference on Machine Learning. pp. 1461–1476. PMLR (2021)
31. Yellin, F., Haeffele, B.D., Roth, S., Vidal, R.: Multi-cell detection and classification using a generative convolutional model. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8953–8961 (2018)