Semi-Automatic Data Annotation guided by Feature Space Projection

Bárbara C. Benato*
Institute of Computing,
University of Campinas,
Campinas, Brazil
Email: barbara.benato@ic.unicamp.br

Alexandru C. Telea
Department of Information and Computing Sciences,
Faculty of Science, Utrecht University,
Utrecht, the Netherlands
Email: a.c.telea@uu.nl

Alexandre X. Falcão
Institute of Computing,
University of Campinas,
Campinas, Brazil
Email: afalcao@ic.unicamp.br

Abstract—Data annotation using visual inspection (supervision) of each training sample can be laborious. Interactive solutions alleviate this by helping experts propagate labels from a few supervised samples to unlabeled ones based solely on the visual analysis of their feature space projection (with no further sample supervision). We present a semi-automatic data annotation approach based on suitable feature space projection and semi-supervised label estimation. We validate our method on the popular MNIST dataset and on images of human intestinal parasites with and without fecal impurities, a large and diverse dataset that makes classification very hard. We evaluate two approaches for semi-supervised learning from the latent and projection spaces, to choose the one that best reduces user annotation effort and also increases classification accuracy on unseen data. Our results demonstrate the added-value of visual analytics tools that combine complementary abilities of humans and machines for more effective machine learning.

I. INTRODUCTION

Machine Learning (ML) models have been extensively investigated and used for regression and classification problems [2]–[4]. More recently, Convolutional Neural Networks (CNNs) have shown great success in many applications, such as image/text classification [5] and speech recognition [6], since they require considerably less effort to optimize parameters than the common feature extraction pipeline [5]. However, CNNs may require a high number of labeled samples (annotated objects) for training [7].

While small labeled training sets can impair the ability of an ML model to correctly classify new samples (a problem known as *over-fitting* [8]), large unlabeled sets make visual inspection and annotation very expensive for the expert. Human costs become even so more prohibitive in domains that require specialized knowledge about the objects, like Medicine and Biology. Solutions for small labeled sets include data augmentation [9] and regularization methods [10]. For large unlabeled sets, semi-supervised classifiers have been used to propagate labels from a small supervised set to the many unsupervised samples by exploring the sample distribution in some feature space [11]–[13]. Yet, none of these approaches has combined the cognitive ability of humans in data abstraction with the ability of machines in data processing to increase the number of labeled objects.

Recent studies have investigated the use of feature space projections and visual analytics to understand and engineer ML models [14]–[18]. Such work addresses both aforementioned labeling cases with approaches for interactive data augmentation [16] and interactive data annotation [17], [18] guided by feature space projections, respectively. Bernard et al. [17] have compared interactive data annotation in a feature space projection with an active learning technique, in which experts supervise and annotate samples selected by a classifier and the classifier is retrained to annotate and select more samples in the original feature space. They discovered that interactive data annotation in the feature space projection is superior to active learning. Benato et. al. [18] have showed that when the user propagates labels to a large unsupervised sample-set guided by the true-label knowledge of a few samples and by the visual information of the sample distribution in a feature space projection, the resulting labeled training-set is more correct than the one created by semi-supervised classifiers in the original feature space. Hence, classifiers trained from such interactively labeled sets can better predict labels of unseen test samples than those trained from automatically labeled sets. Yet, Bernard et al. [17] and Benato et al. [18] have not *combined* automatic and interactive approaches for label propagation — i.e, they have not been concerned with the user *effort* in visual data inspection and annotation.

In this master dissertation, we fill the above gap by proposing a semi-automatic approach that reduces user labeling effort while achieving better classification accuracy on unseen test sets. For this, we exploit the concept of *sample informativeness* from Active Learning (AL). Such approaches select samples for expert supervision based on their informativeness — i.e., potential to improve the design of a classifier from the knowledge of their true label [19], measured by the *confidence* of a classifier about the label assigned to a sample [20]–[23]. In our case, we propagate labels to samples with high-confidence values; and enable the expert focus on low-confidence values for manual label propagation. For this, the user visually analyzes the sample distribution in a 2D scatterplot created by the *t-Distributed Stochastic Neighbor Embedding* (t-SNE) technique [24], constructed similarly to [17], [18], and the true-label knowledge

of only a few samples per class. Although our method can explore further classifier improvement of the classifier by multiple iterations of AL with additional supervised samples, we solve data annotation from a single user interaction for label propagation with no sample supervision. For automatic label propagation, we evaluate two semi-supervised classifiers trained in both latent and projection spaces for automatic label estimation and choose the best one for our goal. We show that our semi-automatic label propagation (SALP) method achieves end-to-end better classification results as compared to both fully automatic label propagation and fully manual label propagation.

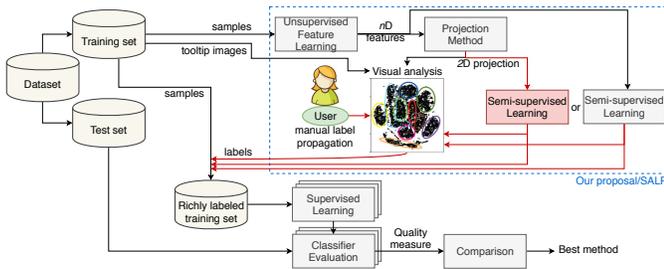## II. SEMI-AUTOMATIC PROJECTION-BASED DATA ANNOTATION

Fig. 1. Semi-automatic data annotation pipeline [25]. We extract features by unsupervised learning from the training set and next use these to project this set to a 2D scatterplot. We next enrich the training set by propagating labels from supervised to unsupervised samples by automatic methods (in both latent and projection spaces) and by manual user-controlled methods. We finally compare the quality of the classifiers trained on such training sets to decide on the best label propagation method. Red indicates additions to earlier related work [18].

Given a training set with a low number of supervised samples and considerably larger number of unsupervised samples, our semi-automatic data annotation approach (Fig. 1) has four steps:

- *unsupervised feature learning:* We start by extracting features from the input dataset. To minimize the number of supervised samples needed, we adopt an unsupervised feature-learning procedure;
- *feature space projection:* We create a feature space 2D projection that captures well the sample distribution in the latent feature space for further visual analysis;
- *semi-supervised label estimation:* We propagate labels automatically to high-confidence unlabeled samples, thereby increasing with training-set size with little effort and high quality;
- *visual analysis:* The expert creates additional labeled samples to the above ones, by interactively propagating labels to the less-confident samples using 2D projection.

A. Unsupervised Feature Learning

We use an Autoencoder Neural Network (AuNN) [26], [27] for unsupervised feature learning. AuNNs consist of two parts, encoder and decoder. The encoder maps the input samples to points in a reduced (latent) feature space; the decoder

reconstructs these samples. The two parts are coupled and trained together by backpropagation. As cost function, we use the mean squared error between the original and reconstructed samples. Hence, we train the AuNN with all labeled and unlabeled samples by ignoring labels. After evaluating several models, we decided for a Stacked Convolutional AuNN [26] — a neural network that presents convolutional layers and can usually obtain relevant latent features.

B. Feature Space Projection

Previous works indicate that 2D projections, created by the t-SNE algorithm [28], [29], achieve this goal well [14], [17], [18], so we follow these (Sec. II-B). The dimension of the latent feature space can still be considered very high (with hundreds to thousands of features) and so unfeasible for visual inspection of the sample distribution. As previously mentioned, we wish to reduce the latent space to two dimensions by preserving as much as possible the relevant structure of the data. The most suitable techniques for this task seem to preserve local distances between samples and the t-SNE algorithm satisfies this criterion [24].

C. Semi-Supervised Label Estimation

For semi-supervised label estimation, we consider two techniques that explore the sample distribution in a given feature space to propagate labels from supervised to unsupervised ones: Laplacian Support Vector Machines (LapSVM) [30], [31] and Semi-Supervised Classification by Optimum-Path Forest (OPF-Semi) [32]. We evaluate both methods on both latent and projection spaces. Given that the performance of OPF-Semi in label propagation is much higher than that of LapSVM (see Sec. III), we select OPF-Semi to output confidence values, used next for our manual label propagation (Sec II-D). Additionally, we found that OPF-Semi in the projection space outperforms itself in the latent feature space (see Sec. III). Hence, we use the 2D version of OPF-Semi for semi-automatic data annotation.

OPF-Semi maps (un)labeled samples to nodes of a graph and computes an optimum-path forest rooted at labeled samples. In this forest, each node s is conquered (labeled) by the root R that offers a path of minimum cost $k(R, s)$ to s. We use costs to compute label confidence values $c(s)$ as described in [21]–[23]. We use the confidence as follows: All labels assigned by OPF-Semi having a confidence above a threshold τ are used as such in the training process. τ is chosen by the user based on the visual analysis of the feature projection with unsupervised samples colored by their confidence values from red (low c) to green (high c) (Fig. 2). Changing τ interactively by a slider lets the user (a) say that high-confidence samples can keep their likely good labels assigned by OPF-Semi and (b) focus on the remaining low-confidence samples to assign them labels by manual label propagation. Users can choose τ balancing how much they wish to trust OPF-Semi *vs* how many samples they are willing to label manually.

D. Manual Label Propagation

The added value of user-driven label propagation in a t-SNE projection was demonstrated by the interactive label propagation technique in [18] which we refer to next as ILP for brevity. However, ILP propagation is fundamentally affected by the quality of the latent features extracted by the AuNN (Sec. II-A) *and* the quality of the t-SNE projection itself: If both these operations faithfully preserve the similarity of original samples, then the user can likely propagate labels well, by simply selecting points close in the projection to the supervised samples. If either the latent space or the projection create errors, which they inherently do [33], this will likely create wrong labels. We assist the user in this process as follows. We color the supervised points in the projection by their labels, and color all low-confidence unsupervised points s having $c(s) < \tau$ in black (Fig. 2(b)). The black points are projected before the colored points, in order to minimize undesired occlusions. When moving the mouse pointer over a projected point, we show its sample image in a tooltip. The user next employs these three sources of information – proximity of unsupervised (black) points in the 2D projection to supervised (colored) ones, low-confidence value of the unsupervised points, and similarity of unsupervised-to-supervised tooltip images – to decide which unsupervised samples get which supervised label. Label propagation is next done simply by selecting desired points in the projection and clicking to assign them a supervised-point label.
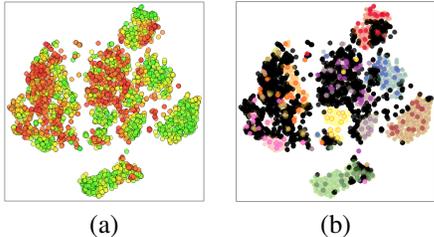

(a)   (b)

Fig. 2. a) Feature projection: unsupervised samples from red (low) to green (high confidence). (b) Semi-automatic label propagation is done from the supervised samples (points colored by class, saturated colors) first automatically to the unsupervised and high-confidence ones (light colors). Low-confident samples (black) are candidates for manual propagation [25].

III. EXPERIMENTS AND RESULTS

A. Experimental Setup

We divide each available dataset D into three subsets for validation: a very small training set S with a few supervised samples per class $(3\%|D|)$; a considerably larger training set U with unsupervised samples for label propagation $(67\%|D|)$; and a set T with unseen test samples $(30\%|D|)$. Next, based on the user-chosen confidence threshold τ, we split U into high-confidence samples L_c, which get their label from OPF-Semi, and low-confidence ones L_i, which can be interactively labeled by the user. Note that $L_c \cap L_i = \emptyset$ and $L_c \cup L_i \neq U$, since the user can choose not to label L_i entirely, to minimize manual labeling effort. We randomly split D into S, U, and T this way three times and repeat the evaluation — i.e., label

propagation from S to U followed by supervised training on $S \cup U$ and testing on T — for statistical purposes.

After labels are propagated from S to U, we train a supervised classifier on $S \cup U$ using the latent feature space. For this task, we used the Optimum-Path Forest (OPF) [34] and Support Vector Machines (SVM) [35]. We test the classifiers on T.

B. Baselines

We propose a semi-automatic label propagation (SALP) that uses OPF-Semi in the 2D t-SNE projection space to propagate labels to high-confidence samples and the user to propagate labels to low-confidence samples, respectively. We next compare SALP with the following three baselines:

1) *No label propagation (NLP):* SVM and OPF, are trained from only S, ignoring set U.
2) *Automatic label propagation (ALP):* set U is fully labeled by one of the four ALP methods below and SVM and OPF are trained from $S \cup U$.
   a) LapSVM using the nD latent feature space.
   b) LapSVM using the 2D t-SNE projection space.
   c) OPF-Semi using the nD latent feature space.
   d) OPF-Semi using the 2D t-SNE projection space.
3) *Interactive label propagation (ILP):* set U is fully labeled by the user and SVM and OPF are trained from $S \cup U$, as in [18].

In all above cases, we test SVM and OPF on T.

C. Datasets

Our first dataset contains 5000 images (28×28 pixels each) of handwritten digits from 0 to 9, randomly selected from the popular public dataset MNIST [36]. Our next three datasets use images (200×200 pixels each) from an automatic processing pipeline that separates microscopy images of human intestinal parasites into three groups: (i) *Helminth larvae* and fecal impurities (3514 images, two classes); (ii) *Helminth eggs* and fecal impurities (5112 images, nine classes); and (iii) *Protozoan cysts* and fecal impurities (9568 images, seven classes). Fecal impurity is a diverse class that has very similar samples to parasites (see Fig. 3). We consider these three datasets with and without images of fecal impurities, yielding five datasets for testing our proposal, apart from MNIST. Those are the most common species of human intestinal parasites in Brazil, which are responsible for public health problems in most tropical countries [37]. All three datasets are unbalanced with considerably more impurity samples.

Fig. 3. Examples of H.Eggs' species (left) and similar impurities (right) [25].

D. Experimental Results

We discuss the performance of our pipeline, measured by the performance of the classifiers trained from $S \cup U$ in the latent feature space and tested on T, by answering the following questions:

- Which space (nD latent, 2D projection) is better for ALP?
- How to set the confidence threshold τ?
- Which approach (manual, semi-automatic, automatic) best propagates labels from S to U?
- What is the end-to-end value of SALP?

Note that we use the 2D projection space only for manual label propagation, i.e. not for testing, since we cannot assume that set T is known during training. Due to the limited space, we opt by omitting the tables and let the graphs with the same results instead.

1) Influence of reducing the feature space from nD to 2D: From the mean and standard deviation of the classification results on set T (Table omitted here, see graphs in Sec. IV), we could get several insights. First, we see that LapSVM performs sometimes better and sometimes worse in nD as compared to 2D, depending on the dataset. In contrast, OPF-Semi consistently shows a positive impact of reducing the feature space independently of the dataset. This happens even when its label-propagation performance is not the best one.

2) The choice of the confidence threshold: As stated in Sec. II-C, users need to choose the threshold τ to specify which automatically-propagated labels they want to keep and which they wish to 'override' manually. Figure 4 shows the projections for the six studied datasets. We see that the threshold τ varies relatively little (0.5 or 0.6) across datasets. This indicates that a good default value to start with is $\tau = 0.5$, after which users can tune τ upwards or downwards depending on the actual distribution of confidences in the projection.

3) Best label propagation approach: From the mean and standard deviation of the classification results on set T (Table omitted here, see graphs in Sec. IV), we showed that OPF-Semi 2D is the winner for automatic label propagation (ALP). Hence, the next question is how well this method would compare against interactive label propagation (ILP) [18], which uses manual label propagation to *all* unsupervised labels, and our new semi-automatic label propagation (SALP), which uses manual label propagation to samples with low-confidence unsupervised labels only. Figure 5 illustrates the ILP and SALP projections for the studied datasets. A key advantage of SALP over ILP is that it shows only the least confident samples (according to OPF-Semi 2D) to the user, hence reducing the effort needed to understand the picture (and also reducing clutter and overlap in the projection), thus making the interactive labeling task easier. We discuss next several observations relating ILP to SALP in Fig. 5.

For the MNIST dataset, the user propagated labels to 1864 unsupervised samples on average (over the three considered runs) when using ILP. When using SALP, this number dropped to 1182 samples. This pattern of less effort for SALP is consistent over all other datasets. We focused our discussion
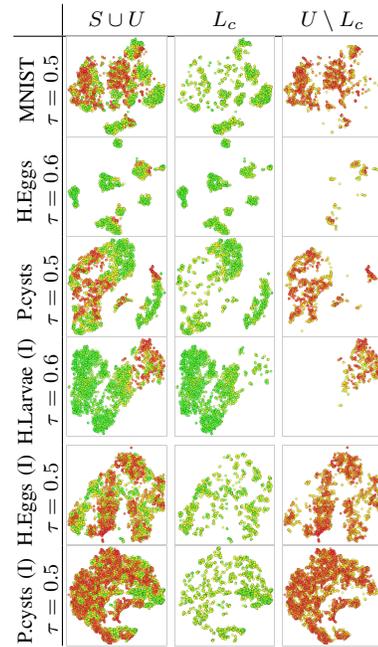

Fig. 4. Projections colored by label confidence (red=low confident, green=high confident). Rows are datasets (easiest at top, hardest at bottom). Columns show the entire set of supervised-and-unsupervised samples $S \cup U$, the high-confidence samples L_c labeled by ALP, and the low-confidence samples $U \setminus L_c$ that go to manual labeling [25].

here in the most difficult datasets: H.Eggs and P.cysts with impurities.

For H.Eggs dataset with impurity, the supervised impurity samples (gray) fall between groups of colored points (actual H.Eggs classes) in the projection. In contrast to the earlier datasets, we see many more black points in SALP, meaning that OPF-Semi 2D has difficulties in automatically propagating labels. This matches the fact that datasets with impurities are considerably harder. For this dataset, the user propagated labels to more points in SALP (2076) than ILP (1787). This seems to support the evidence that the simplification of the SALP projection by removing high-confidence points, even though minor in this case, was enough to help the user see more structure in the projection along which she could propagate labels. Also, as for P.cysts, we see that OPF-Semi 2D propagates labels in more central regions of visible groups, leaving the rest to the user. Finally, for P.cysts with impurities, the supervised impurity samples (brown) are spread out over the entire projection. The supervised P.cysts samples (other colors than brown) are mixed quite strongly, and the projection shows little structure – roughly, one large and one small crescent-shaped group. This is the most challenging dataset for manual label propagation. When adding the impurities to those datasets, the difficulty increases for the classifiers (Sec. III-D4).

As for H.Eggs, OPF-Semi 2D finds only few confident samples, so the manual labeling effort is quite similar for both ILP and SALP. This is matched by the actual number of points to which the user actually propagated labels (1787 with ILP
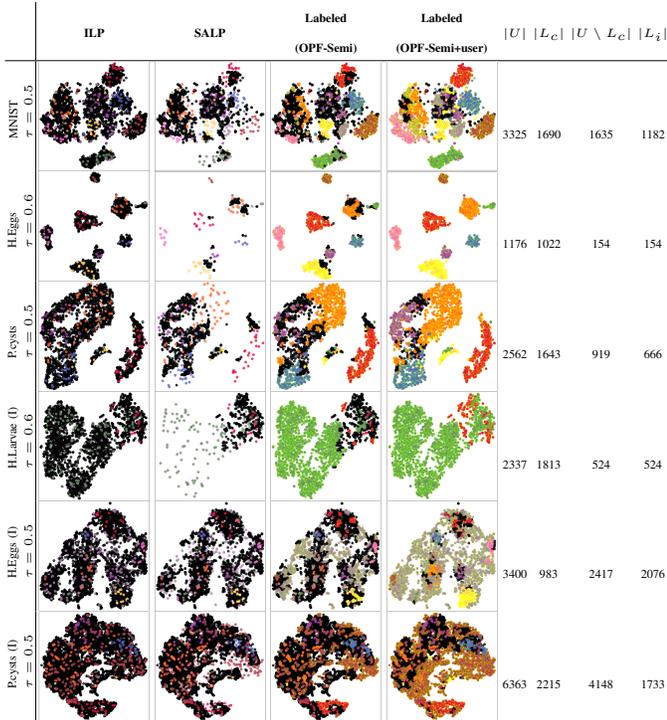
Fig. 5. Comparison of different label propagation methods (columns) for different datasets (rows). From left to right: ILP, SALP, labels automatically propagated by OPF-Semi, and final labeling result of SALP together with OPF-Semi. Colors indicate labels given by either supervised samples (ILP, SALP) or both unsupervised and propagated labels (OPF-Semi, OPF-semi+user)). Black shows samples to be considered by manual propagation (three left columns), and samples skipped by manual propagation (right column). Sample set sizes are shown to the right [25].

vs 1733 with SALP). Even though these figures are almost identical, the main benefit for SALP here is that OPF-Semi 2D already filtered the easy cases (high confidence) points, thereby focusing the user's effort to the more difficult cases.

4) End-to-end value of SALP: We have seen that SALP decreases the user's effort in label propagation. A final question we answer is: How much added-value does SALP bring, in terms of classification quality, as opposed to the earlier similar method, ILP, or to the best fully-automatic counterpart we found, OPF-Semi 2D? From the mean and standard deviation of the classification results on set T (Table omitted here, see graphs in Sec. IV), we see that SALP consistently obtained the best classification results on unseen T for all datasets. This proves that SALP is, indeed, of added value with respect to earlier existing methods – using it yields better classifiers in the end. Separately, we see that, for all but the simplest datasets (MNIST and H.Eggs), SALP also yields the best label propagation accuracy.

IV. DISCUSSION

A. Using the nD vs 2D feature space

An interesting question is how the fully automatic label propagation (ALP) performs when using the latent nD feature space *vs* the 2D projection space. Figure 6 shows the average κ classification values for LapSVM and OPF-Semi using these

two spaces for the OPF and SVM classifiers respectively. Datasets are sorted along the x axis by decreasing order of the κ value for OPF-Semi 2D. We see that LapSVM leads to better results in 2D than in nD for half of the datasets, while OPF-Semi does that for *all* datasets. This essentially tells that the 2D projection space, created by t-SNE, is able to retain all needed information to enable the desired label propagation and, next, good-quality classifier construction. This is an important result, as it justifies next presenting the 2D projection space to the user as the sole information based on which she will perform the manual label propagation. We also see that the trend of the κ values along the x axis, for both the 2D and nD variants, matches the perceived difficulty of the datasets: High κ values correspond to easier datasets (left), while lower κ values correspond to the harder datasets with impurities (to the right). Finally, we plot here also the κ values for ILP and SALP (curves in the figures). In all cases, these curves are above the automatic methods, showing that adding manual effort pays off. The SALP curve is above the ILP one, showing that the optimal design is reached by *combining* automatic and manual propagation (both in 2D).
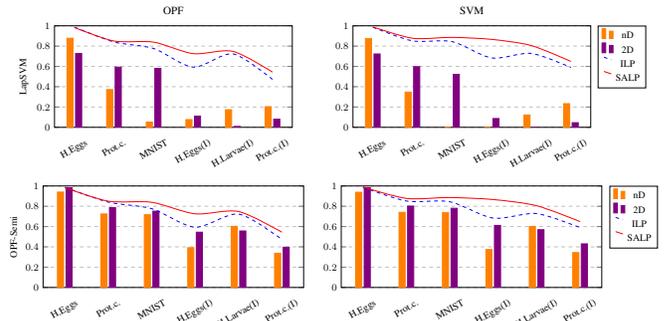

Fig. 6. κ values for studied datasets, for OPF and SVM classifiers (columns) and LapSVM and OPF-Semi automatic label propagation methods (rows). Curves show κ values for ILP and SALP. Datasets are sorted from easiest to hardest to classify (left to right) based on SALP results [25].

B. User effort reduction

Besides achieving the best classification results, as compared to both fully-automatic and fully-manual (ILP) label propagation, SALP also reduces the *manual* effort as compared to ILP. Figure 7 shows this by depicting the percentage of samples labeled by the user over total number of samples to label ($|U|$) per dataset and for ILP and SALP. For SALP, this measurement excludes, indeed, the automatically-labeled samples by OPF-Semi 2D. Datasets are sorted along x by increasing $|U|$, i.e, from the smallest to the largest dataset. Figure 7 reveals several insights. First, assuming that the labeling effort is proportional with the number of labeled samples and the effort per sample is the same for ILP and SALP (which should be the case given that the two methods share the same visualization and interaction), we see that the ILP effort is always larger than the SALP effort, except for H.Eggs with impurities. Secondly, the percentage of propagated samples for ILP decreases with the dataset size. This can be explained

by the difficulty of propagating labels in projections showing many points, where overlap and clutter become issues. We note an opposite for trend SALP: The percentage of propagated samples increases with dataset size. The trend breaks for the largest dataset (Prot.c.(I), 6363 samples), about twice larger than the second-largest dataset (H.Eggs(I), 3400 samples). Here, the projection is likely quite dense and cluttered, so manual propagation becomes similarly hard for ILP and SALP.

In parallel, we observe that the number of samples $U \setminus L_c$, those above the threshold τ and low-confidence labels to OPF-Semi, also increases with the dataset size. Thus, the amount of samples $U \setminus L_c$ presented to the user to propagate labels with SALP increases with dataset size. One case in point is the H.Eggs with impurities dataset. This dataset has the largest percentage of annotated samples by SALP, exceeding also ILP. This is explained by the size of the dataset (second largest one) and the fact that its projection makes it reasonably easy to propagate labels for the large impurity class (Fig. 4).
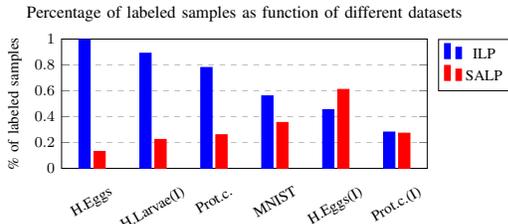

Fig. 7. Percentage of labeled samples in U *vs* ILP and SALP label propagation methods. x-axis is sorted by the number of samples in the datasets [25].

C. Effectiveness

As shown in Fig. 6, SALP consistently yields best classification results, for both SVM and OPF classifiers, overpassing fully manual propagation methods (ILP) and the best fully automatic one (OPF-Semi 2D). The gains of SALP are higher for the more challenging datasets, where fully automatic methods encounter challenges. Conversely, where such methods work well, they reduce user effort as compared to fully manual propagation (ILP). In brief, this shows that the combination of automatic methods with human insights is indeed of added value both in increasing classifier quality and decreasing the effort needed to achieve it.

D. Limitations

First, *validation* is limited to six datasets, two classifier techniques, and one user performing manual labeling. Measuring the added-value of SALP for more (dataset, classifier, user) combinations would bring more insights into the effectiveness of the method. Secondly, while the added-value of the 2D t-SNE projection space has been demonstrated, the actual effect of t-SNE's distortions has not been quantitatively gauged. Using projection accuracy metrics such as stress, trustworthiness, continuity, or neighborhood hit [33] can be used to find such correlations. On the other hand, using visual tools [33] that highlight such errors can help the user to achieve more accurate and/or faster manual label propagation.

V. Conclusion

We proposed a combined automatic-and-user-driven approach for creating labeled samples for sparsely-annotated datasets for the purpose of training classifier models. For this, we extract dataset features using Autoencoder Neural Networks and next reduce these to a 2D space using t-SNE. We next automatically propagate labels from the (few) supervised to unsupervised samples in this 2D space, while monitoring the propagation confidence. For low confidence labeled samples, we allow the user to manually annotate them by using the visual insights encoded in the 2D projection annotated with the supervised sample labels. Several quantitative results follow: First, we showed that the 2D projection space leads to higher-accuracy automatic label propagation than the high-dimensional latent space extracted by the autoencoder. To our knowledge, this insight is new, and suggests new ways for dimensionality reduction. Secondly, we show that our semi-supervised method, combining the OPF-Semi automatic label propagation with user-driven manual label propagation, both done in the 2D space, achieves higher classification quality than both fully-automatic and fully-manual label propagation. This opens the way to different methods for combining automatic and human-centered methods for the engineering of high-quality machine learning systems.

Future work will consider the use of the proposed semi-automatic label propagation method in Active Learning (AL) scenarios. We expect that AL looping can improve classification results. Separately, we plan to perform more extensive validation studies measuring the added-value of our approach for more types of datasets, classification methods, and using additional visual analytics techniques to help users to propagate labels better and faster.

The following distinctions resulted of this dissertation:

- Three international conference papers [16], [18], [38];
- One international journal paper at *Pattern Recognition* [25];
- Best poster award by the entitle work "Data annotation guided by Feature Projections", XIV Workshop de Teses, Dissertações e Trabalhos de Iniciação Científica (WTD), IC UNICAMP, 2019;
- Six months research internship at University of Groningen funded by BEPE/Fapesp and supervised by Prof. Dr. Alexandru Telea.

References

[1] B. Benato, "Anotação de dados guiada por projeção de características," Master's thesis, Universidade Estadual de Campinas, Instituto de Computação, Campinas, SP., 2019.

[2] A. Krizhevsky, I. Sutskever, and H. E. Geoffrey, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[3] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[4] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 1334–1373, Jan. 2016.

[5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.

[6] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, Nov 2012.

[7] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS'14. Cambridge, MA, USA: MIT Press, 2014, pp. 3320–3328.

[8] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.

[9] R. Mash, B. Borghetti, and J. Pecarina, "Improved aircraft recognition for aerial refueling through data augmentation in convolutional neural networks," in *Proc. ISVC*. Springer, 2016, pp. 113–122.

[10] S. J. Nowlan and G. E. Hinton, "Simplifying neural networks by soft weight-sharing," *Neural Computation*, vol. 4, no. 4, pp. 473–493, 1992.

[11] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling, "Semi-supervised learning with deep generative models," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 3581–3589.

[12] G. Forestier and C. Wemmert, "Semi-supervised learning using multiple clusterings with limited labeled data," *Information Sciences*, vol. 361–362, pp. 48–65, 2016.

[13] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, and K. Talwar, "Semi-supervised knowledge transfer for deep learning from private training data," in *Proceedings of the International Conference on Learning Representations*, 2017.

[14] P. Rauber, A. Falcão, and A. Telea, "Projections as visual aids for classification system design," *Information Visualization*, 2017.

[15] P. E. Rauber, A. X. Falcão, and A. C. Telea, "Visualizing time-dependent data using dynamic t-SNE," in *Proceedings of the Eurographics / IEEE VGTC Conference on Visualization: Short Papers*, ser. EuroVis '16, 2016, pp. 73–77.

[16] A. Z. Peixinho, B. C. Benato, L. G. Nonato, and A. X. Falcão, "Delaunay triangulation data augmentation guided by visual analytics for deep learning," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2018, pp. 384–391.

[17] J. Bernard, M. Hutter, M. Zeppelzauer, D. Fellner, and M. Sedlmair, "Comparing visual-interactive labeling with active learning: An experimental study," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 298–308, Jan 2018.

[18] B. C. Benato, A. C. Telea, and A. X. Falcão, "Semi-supervised learning with interactive label propagation guided by feature space projections," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2018, pp. 392–399.

[19] B. Settles, "Active learning literature survey," University of Wisconsin–Madison, Computer Sciences Technical Report 1648, 2009.

[20] S. Patra and L. Bruzzone, "A batch-mode active learning technique based on multiple uncertainty for SVM classifier," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 3, pp. 497–501, May 2012.

[21] P. A. V. Miranda and A. X. Falcão, "Links between image segmentation based on optimum-path forest and minimum cut in graph," *Journal of Mathematical Imaging and Vision*, vol. 35, no. 2, pp. 128–142, Oct 2009.

[22] T. V. Spina, P. A. V. Miranda, and A. X. Falcão, "Intelligent understanding of user interaction in image segmentation," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 26, no. 02, p. 1265001, 2012.

[23] A. T. Silva, J. A. Santos, A. X. Falcão, R. S. Torres, and L. P. Magalhães, "Incorporating multiple distance spaces in optimum-path forest classification to improve feedback-based learning," *Computer Vision and Image Understanding*, vol. 116, no. 4, pp. 510 – 523, 2012.

[24] L. V. D. Maaten, "Accelerating t-SNE using tree-based algorithms," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3221–3245, 2014.

[25] B. C. Benato, J. F. Gomes, A. C. Telea, and A. X. Falcão, "Semi-automatic data annotation guided by feature space projection," *Pattern Recognition*, vol. 109, p. 107612, 2021.

[26] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. Intl. Conf. on Artificial Neural Networks (ICANN)*. Springer, 2011, pp. 52–59.

[27] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, Dec. 2010.

[28] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[29] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[30] V. Sindhwani, P. Niyogi, and M. Belkin, "Beyond the point cloud: From transductive to semi-supervised learning," in *Proceedings of the 22Nd International Conference on Machine Learning*, ser. ICML '05. New York, NY, USA: ACM, 2005, pp. 824–831.

[31] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Dec. 2006.

[32] W. P. Amorim, A. X. Falcão, J. P. Papa, and M. H. Carvalho, "Improving semi-supervised learning through optimum connectivity," *Pattern Recogn.*, vol. 60, no. C, pp. 72–85, Dec. 2016.

[33] L. Nonato and M. Aupetit, "Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment," *IEEE TVCG*, 2018.

[34] J. P. Papa, A. X. Falcão, V. H. C. Albuquerque, and J. M. R. S. Tavares, "Efficient supervised optimum-path forest classification for large datasets," *Pattern Recognition*, vol. 45, no. 1, pp. 512 – 520, 2012.

[35] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, July 1998.

[36] Y. LeCun and C. Cortes, "MNIST handwritten digit database," 2010. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[37] C. T. N. Suzuki, J. F. Gomes, A. X. Falcão, S. H. Shimizu, and J. P. Papa, "Automated diagnosis of human intestinal parasites using optical microscopy images," in *2013 IEEE 10th International Symposium on Biomedical Imaging*, April 2013, pp. 460–463.

[38] S. B. Martins, B. C. Benato, B. F. Silva, C. L. Yasuda, and A. X. Falcão, "Modeling normal brain asymmetry in MR images applied to anomaly detection without segmentation and data annotation," in *Medical Imaging 2019: Computer-Aided Diagnosis*, K. Mori and H. K. Hahn, Eds., vol. 10950, International Society for Optics and Photonics. SPIE, 2019, pp. 71 – 80. [Online]. Available: https://doi.org/10.1117/12.2512873