Active Class Selection for Dataset Acquisition in Sign Language Recognition

Manuele Bicego[1] (Γ), Manuel Vázquez-Enríquez[2], and José L. Alba-Castro[2]

[1] Computer Science Dept., University of Verona, Verona, Italy,
manuele.bicego@univr.it
[2] atlanTTic research center, University of Vigo, Vigo, Spain,
{mvazquez,jalba}@gts.uvigo.es

Abstract. Dataset collection for Sign Language Recognition (SLR) represents a challenging and crucial step in the development of modern automatic SLR systems. Typical acquisition protocols do not follow specific strategies, simply trying to gather equally represented classes. In this paper we provide some empirical evidences that alternative, more clever, strategies can be really beneficial, leading to a better performance of classification systems. In particular, we investigate the exploitation of ideas and tools of Active Class Selection (ACS), a peculiar Active Learning (AL) context specifically devoted to scenarios in which new data is labelled *at the same time* it is generated. In particular, differently from standard AL where a strategy asks for a specific label from an available set of unlabelled data, ACS strategies define *from which class* it is more convenient to acquire a new sample. In this paper, we show the beneficial effect of these methods in the SLR scenario, where these concepts have never been investigated. We studied both standard and novel ACS approaches, with experiments based on a challenging dataset recently collected for an ECCV challenge. We also preliminary investigate other possible exploitations of ACS ideas, for example to select which would be, for the classification system, the most beneficial *signer*.

Keywords: Active Class Selection · Sign Language Recognition · Active Learning.

1 Introduction

Automatic Sign Language Recognition (SLR) represents a classic Pattern Recognition problem, which interest has increased in recent years due to the latest advances in deep learning models with flexible spatial-temporal representation capacities [18]. SLR can be broadly categorized into Isolated (ISLR) and Continuous (CSLR) sign language recognition. ISLR is the most extensively researched scenario, and more annotated datasets are available for it due to the relatively simple, but tedious, annotation protocols of a discrete and predefined set of signs. On the other hand, CSLR is much more complex, and much fewer annotated datasets exist. This is due to different challenges related to annotation: i) co-articulation between signs significantly increases the variability of the sign realization with respect to isolated signs; ii) the speed is much higher, and iii) there is an interplay of non-manual components such as torso, head, eyebrows, eyes, mouth, lips, and even tongue [8].

Still, it is widely acknowledged that proper SLR systems require massive amounts of data [18]. In typical scenarios, datasets are collected so that all possible class instances are equally represented, and performed by all the available experts (the signers); however, in some contexts, this is not a doable strategy, sometimes because of restrictions in the availability of signers, sometimes because signs are simply gathered from real-world videos [2, 22]. The definition of a proper strategy to acquire data is therefore crucial, and deserves more attention than that which is typically devoted to it: actually, the standard solution is to simply try to gather equally represented classes, labelled by all signers. However, it is highly possible that having a non-homogeneous distribution of classes is more useful for the classifier, since it is quite normal that some signs have more intrinsic or per-signer variability, and they are more complicated for the classification. Moreover, also the number of samples which are assessed by each signer should be carefully decided: for example, in the distributed Isolated Sign Language Recognition (ISLR) acquisition system described in [23], the collaboration of deaf individuals is voluntary, so it is crucial to minimize the number of instances required per sign to prevent signers from becoming disinterested in contributing their time.

Given all these observations, it seems evident that it would be very beneficial to define a proper dataset construction strategy: and this aspect, never considered in the SLR scenario, represents the main goal of this paper. In particular, here we investigate the usefulness of a particular class of Active Learning strategies [20], called Active Class Selection (ACS – [15]), for the construction of an SLR dataset. ACS represents a very specific Active Learning paradigm, introduced for a very peculiar scenario, which goes beyond the classic Active Learning assumption that getting objects is cheap, whereas getting labels is not. Actually, in ACS the assumption is that labels are directly obtained when getting data, since the experiments are designed *given the labels*. The example, reported in [15], is odour classification with an electronic nose. In a typical setup, the odorant is let to flow over the array of sensors, from which the signal is recorded. In this scenario, the label is decided in advance (the odorant), and the recording is acquired together with the label. Therefore, differently from standard AL, where we have at disposal a set of unlabelled data, and the AL strategy asks for a specific label, ACS strategies aim at selecting the best "class" from which we should get a sample, without having unlabelled data. ACS strategies have not been largely studied, due to their very specific applicability: examples can be found in the already cited odour classification [19, 15], Brain-Computer Interfaces [17, 24, 25], geology [14] and robotics [6]. Also, from a theoretical and methodological point of view, Active Class Selection has not been largely investigated, with an increasing interest only in recent years, with the work of Kottke and colleagues [13], and, especially, with the very recent theoretical works of Bunse and colleagues [5, 4].

In this paper, we investigate for the first time the usefulness of ACS techniques in the SLR context: in particular, we show that ACS strategies can be very beneficial in this context, and can improve the classic option of simply getting equally numbered classes. We tested different ACS options, also proposing a novel computationally cheaper version based on Random Forests. We designed a pipeline to test all strategies, evaluating them with a real-world dataset [22], in which classification is based on a pipeline defined with Graph Convolutional Networks [21], showing that ACS strategies can be very useful in the SLR context. We also preliminary study the possible exploitation of the ACS ideas and strategies to face another aspect: the selection of the signers. Signers are very different, can be true deaf people or sign-language interpreters. Differences are large, and selecting the "best" signer for a given sample can be possibly very important. Results, in this case, are mixed, and further investigations are needed.

2 Sign Language Recognition

2.1 Dataset creation for SLR

The acquisition of 3D spatial-temporal information from sign-language gestures has traditionally involved a multitude of devices such as motion-capturing systems [12], depth-based sensors [1], ultrasound-based sensors [7], and even wearable sensors [26]. Thus, the collection of samples required lab facilities and a slow and costly process of recording deaf signers one by one [10]. However, recent research has shown a trend towards the utilization of only RGB inputs, without any supplementary capturing device. With the emergence of Deep Learning methods, accurate depth estimation can be achieved through learned body models [11], so data acquisition with cumbersome devices can be avoided. Raw RGB-based approaches typically employ Convolutional Neural Network (CNN)-based backbones for spatial feature extraction and Recurrent Neural Network (RNN) for temporal encoding or a spatio-temporal feature extractor (3DCNN) commonly used in Human Action Recognition (HAR) tasks. Current research suggests that spatio-temporal graph convolutional networks (ST-GCN) [27] is the state-of-the-art approach for this purpose.

The class samples used for this research on ACS are extracted from the LSE_eSaude_UVIGO dataset used for the 2022 Sign Spotting Challenge at ECCV'22 [22]. Each sign instance annotated in the sign language dataset contains a variable number of frames. The duration of the co-articulated signs starts from 120 ms (3 frames) up to 2 sec (50 frames), with a mean duration of 520 ms (13 frames). We will provide more specific information in the experimental part.

2.2 A DL pipeline for SLR

The ACS scheme can be seen as a procedure that defines from which class the next object has to come. In our study, we aim at assessing the potential classification improvements due to such careful selection of classes: in the following, we describe the Deep Learning pipeline which we used as basis for computing classification accuracy. This pipeline, summarized in Figure 1, represents the first part of the inference scheme provided as the baseline of the cited Challenge, and it is explained in more details in section 3.3 of [22]. The pipeline takes a sliding temporal window of 400 ms containing the whole sign instance (or a part of it) and outputs an embedding vector with spatial-temporal information. The embedding vector is collected at the output of an ST-GCN network and just before the softmax layer; each annotated sign instance is then represented with a variable number of embedding vectors (of dimension 60) that span from 400 ms to 2 sec of spatial-temporal sign information. More formally, each sign x in

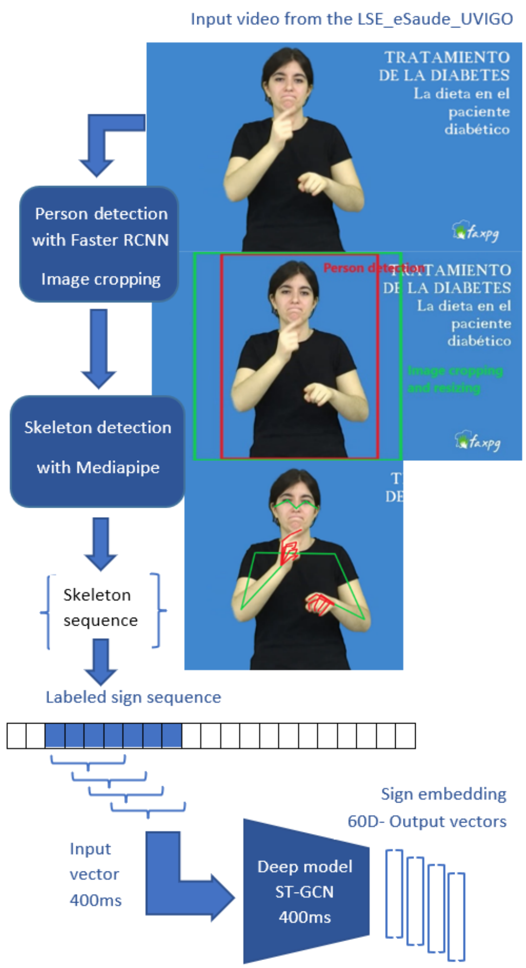


Fig. 1. Pipeline to process an RGB sign instance and convert it to a sequence of embedding spatial-temporal vectors.

the dataset X is a collection of 60-dimensional vectors (the number of vectors in x depends on the length of the sign x), i.e.

$$x_k = [x_k^1, x_k^2, \cdots, x_k^l], \qquad x_k^i \in \Re^{60}$$

We call each vector x_k^j as *frame*, even if it derives from the analysis of a subsequence of 400 msec of the original video, as explained above. To perform classification, instead of using the procedure described for the Challenge baseline [22] (which involves a second computationally demanding ST-GCN), we investigate a simpler approach based on a Multiple Instance Learning scheme, a recent learning paradigm [9] which extends classical supervised learning. In this paradigm, each object is represented by an unordered set of feature vectors, called instances. This set of instances, called a bag, has a unique label. By considering the frames inside a sign as the instances, and the sign as a bag, we can apply this learning paradigm, whose usefulness has been assessed in many different fields, but not (yet) in the sign language recognition scenario. Of course, we are losing the order in which the different frames appear, but, in many applications, this restriction still permits us to have excellent performances [16]. In particular, here we used simple bag-based approaches, methods which summarize each bag with a single feature vector: more in detail we used the Max pool strategy: in this scheme, the sign $x = [x^1, x^2, \cdots, x^l]$ is summarized with a single vector z, in which each entry $z(h)$ is just the max over all the values of the entry h of the vectors $x^1, x^2, \cdots, x^l$:

$$z = [z(1), z(2), \cdots z(60)]$$

where

$$z(h) = \max_{d=1,\ldots,l} x^d(h)$$

Given this representation, we used as classifier the Random Forest Classifier [3], with 100 trees.

3 The proposed approach

3.1 Active Class Selection

As described above, the ACS scheme can be seen as a procedure that defines from which class the next object has to come. Here we adopt the formulation used in [15], which defined a batch-based iterative procedure to be used to define a sequence of training sets $X_0^{tr}, X_1^{tr}, \cdots, X_i^{tr}, \cdots$ of increasing size $N_0, N_0+N, N_0+2N \cdots$, in which, given the whole set of possible examples X and a starting training set X_0^{tr}:

- each X_i^{tr} is a subset of X ($X_i^{tr} \subset X$), $|X_0^{tr}| = N_0$, $|X_i^{tr}| = N_0 + iN$, where $|\cdot|$ denotes the cardinality of a set.
- X_i^{tr} includes X_{i-1}^{tr} ($X_{i-1}^{tr} \subset X_i^{tr}$). In other words, each dataset X_i^{tr} is obtained by adding to X_{i-1}^{tr} N elements sampled from $X \setminus X_{i-1}^{tr}$.

The **Active Class Selection** (ACS) strategy defines how to get X_i^{tr} from X_{i-1}^{tr}, i.e. defines how to sample N new objects. In particular, every ACS strategy defines the proportion of classes at time i, i.e. which classes deserve more objects and which less. This is typically formalized as a multinomial $\mathbf{p}^i = [p_1^i, p_2^i, \cdots p_C^i]$, where C is the number of classes and $\sum_c p_c^i = 1$. To define this multinomial $\mathbf{p}^i$ we can use all the samples in X_{i-1}^{tr}; the new objects are sampled from $X \setminus X_{i-1}^{tr}$ according to $\mathbf{p}^i$: in particular, if we have to sample N objects, for each of them we select a class according to $\mathbf{p}^i$, taking one random object from that class.

3.2 The ACS strategies

In the paper we implemented different strategies, some already presented in [15] and some others adapted for the specific context.

Baseline-Random: this represents the baseline and the usual way of creating the dataset, where all classes are equiprobable (i.e. we try to have all classes with the same number of signs). In this case $\mathbf{p}^i = [1/C, 1/C, \cdots 1/C]$.

Baseline-Natural: in this case, $\mathbf{p}^i$ reflects the *true* distribution of the classes of the problem, which is typically unknown. In our application, however, the signs are extracted from real life video sequences, thus we can estimate it by checking the composition of the whole dataset X.

ACS-Inverse: in this case, each p_c^i is defined as inversely proportional to the accuracy of the classifier on the class c. The idea is that if a class is poorly classified, then we need more samples for that class. We compute the accuracy of the classifier described in previous section on each of the classes of the set X_{i-1}^{tr}, using a 5-fold cross validation protocol.

ACS-AccImprovement: this is the Accuracy improvement method proposed in [15], in which the idea is to give less importance to classes for which there was not a significant improvement in accuracy in the previous iteration. The idea is that we do not need objects for classes which did not benefit from additional samples in the previous iteration.

ACS-Redistricting: this is the Redistricting method proposed in [15], in which the idea is to ask more sample for classes which are more "unstable", i.e. for which there has been a large change in label assignments in the previous iteration.

ACS-RF-impurity. This represents a novel ACS strategy which we propose here, in order to have a computationally simpler approach with respect to ACS-Inverse. The main idea is that we can learn the proportions $\mathbf{p}^i$ by observing the posterior probabilities of a Random Forest, exploiting the usual out-of-bag mechanism. In this way we are not required to perform the whole classification task (as in the ACS-Inverse), but we can extract $\mathbf{p}^i$ from a single trained RF. In particular:

- we train a single RF on the whole dataset;
- each object in the out-of-bag sample (i.e. the objects not used for training a tree) is let fall down each tree and classified by assigning it to the class

most probable in the leaf where it arrives. The probability of each class, at leaf level, is determined during training, and is basically proportional to the number of training objects of each class falling in such leaf. The posterior of the object x, with respect to class k, is denoted as $\mathrm{Prob}(k|x)$.
- the object of class c would contribute to the uncertainty of class c with the factor
$$U(x) = 1 - \sum_{k \neq c} \mathrm{Prob}(k|x) \tag{1}$$
- the total uncertainty of class c is obtained by averaging $U(x)$ over the objects of class c;
- $\mathbf{p}^i$ is proportional to uncertainty: the more uncertain a class, the larger the number of objects which are required.

4 Experimental Evaluation

In our experiments, we used part of the dataset LSE_eSaude_UVIGO from the 2022 Sign Spotting Challenge at ECCV'22 [22]. This dataset has 100 signs located within 10 hours of Spanish sign language videos. The challenge had two different Tracks, and we focus here on the Track1. The number of instances per sign is largely not uniform, so we selected a subset of signs with at least 30 samples. We create two different versions, each one including 28 different classes: in the first (**ACSDataset1**), for each class we extract 30 random samples, resulting in 840 samples. This represents a dataset with equally represented classes. In the second, (**ACSDataset2**), for each class we kept all samples, resulting in a total of 1639 samples. The smallest class ("EPOCA") contains 30 samples, the largest ("ENFERMO") 176.

To assess the usefulness of the ACS methods we used the protocol proposed in [15], which, starting from a given dataset X, simulates an ACS scenario. In particular, we split the dataset X in two random parts: X^{tr}, used for training and X^{val}, used for validation. The goal is to compute the classification accuracy on X^{val} of the classifier described in the previous section, trained using training sets of increasing size $X_0^{tr}, X_1^{tr}, \cdots, X_i^{tr}, \cdots$ defined in section 3. By plotting the accuracy obtained for the different training sets, we have the ACS curve. More precisely, in our experiments we perform a random 50%-50% splitting of the dataset X – the random split is so that the class proportions in X are maintained in X^{tr} and X^{val}. The initial set X_0^{tr} contains 100 elements uniformly sampled from X^{tr}, whereas at every iteration we added 50 samples. For each of them, i) we select a class, according to the $\mathbf{p}^i$ given by the ACS strategy, and ii) we take one random object from that class. If in the dataset the selected class does not have any further object, we select another random class from those classes which still have objects (this is the same scheme adopted in [15] and subsequent papers). Due to this, the last part of the ACS curve is the less representative. In order to avoid random fluctuations, the whole procedure is repeated 100 times, and results are averaged.

4.1 Results

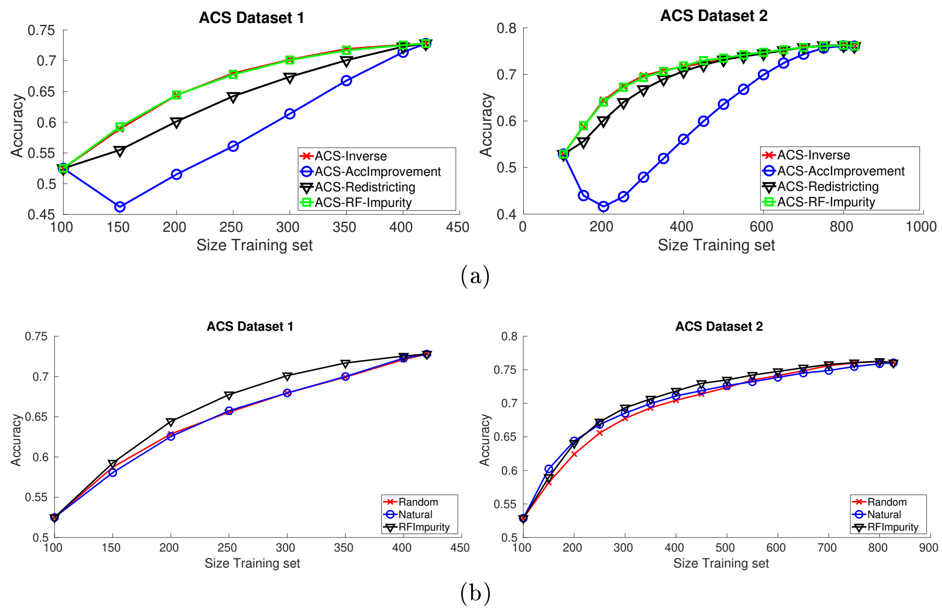


Fig. 2. Active Class Selection results on the two datasets: (a) comparison between different ACS strategies; (b) comparison between the RF-impurity ACS strategy and conventional approaches.

The results are shown in Fig. 2. In particular, we performed two analyses: in the former, reported in part (a) of the figure, we compare the 4 different ACS strategies, for the two datasets. From the plot it seems evident that the ACS-Inverse and the ACS-RF-impurity represent the best option, for both datasets[3]. It is important to note that ACS-RF-impurity represents a computationally cheaper version, with respect to the ACS-Inverse one, reaching a comparable level of accuracies. In order to get an idea of this equivalence, we performed a statistical analysis: in particular, for each training set size, we made a t-test on the 100 repetitions of the two methods, checking if the two populations are different in a statistically significant way (threshold 0.05): for all sizes of training set the two methods are equivalent, with one exception: training set size equal to 450 for Dataset 2. This confirms that the lighter ACS-RF-impurity represents a viable alternative to the ACS-Inverse scheme. In the second part of the plot (part (b)), we compare our proposed ACS strategy (ACS-RF-impurity) with the two conventional alternatives of “Baseline-Random” and “Baseline-Natural”. Also in

[3] We are still trying to understand the strange behaviour of the ACS-AccImprovement strategy, which at the very beginning decreases the performances.

this case we performed a statistical analysis: after a Bonferroni correction for multiple tests, we observed that ACS-RF-impurity is better than the Baseline-Random option with a statistical significance for all training set sizes, except size 150, and than the Baseline-Natural option with a statistical significance for all training set sizes except size 400. For dataset 2, ACS-RF-impurity is better than the Baseline-Random option with a statistical significance for all training set sizes except the first and the last 3, and better than the Baseline-Natural option with a statistical significance for all training set sizes after the first three. Please note that due to the protocol, the most significant part is the central one, since at the end there is no choice, and at the very beginning they are all the same. From these results it seems evident that a proper strategy for constructing the dataset, as that suggested by ACS strategies, can be very beneficial for the SLR scenario.

5 A preliminary investigation on Active *Signer* Selection

In the previous section, we provided some evidences that properly selecting the class of novel samples can be beneficial for the classification system. Here we provide some preliminary experiments on a related aspect: can we expect improvements by properly selecting the *signer*? Actually, signing styles show a large inter-signer variation, even larger than speaking styles – e.g. Tables 5 and 6 of [22] show this effect on the F1-score of Track1 for all the classification systems compared. To investigate this aspect we devise a set of strategies, which we call Active Signer Selection (ASS) strategies, which are very similar to ACS ones, but do not select a class, but a signer. Even if we can also think to perform selection of the signer *and* the class, here we investigate a simpler option: once the signer is selected the class is then selected random from signs of that signer. The strategies we investigated are:

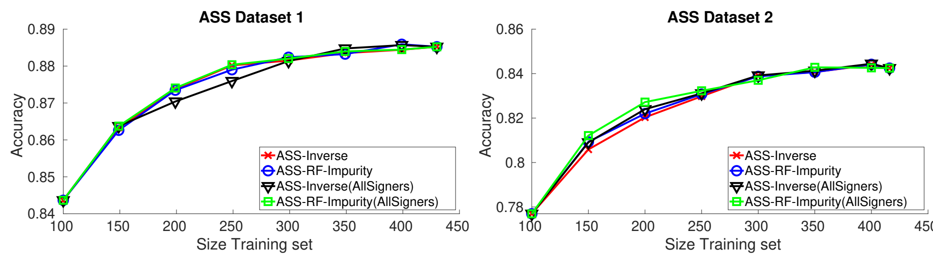
Baseline-RandomSigner, Baseline-NaturalSigner: the baselines, in which we select the signer randomly (Baseline-RandomSigner) or according to the natural proportions of signers (Baseline-NaturalSigner).

ASS-Inverse: inspired from the ACS-Inverse strategy, we give a higher probability to be extracted to the signer who is performing worst. To compute the performance, we select in X^{tr}_{i-1} all the signs scored by a given signer, computing how good is the classifier on such dataset. We also implemented another version, which we called **ASS-Inverse(AllSigners)**, in which we created a single classifier, dividing then the errors signer by signer. This seems to be more adequate when X^{tr}_{i-1} is small.

ASS-RF-Impurity: similar to ACS-RF-Impurtity, but now Random Forest uncertainties are computed and aggregated signer by signer. Also in this case we tested another version, called **ASS-RF-Impurity(AllSigners)**, which implements the same modification discussed for ASS-Inverse(AllSigners).

5.1 ASS results

To test the strategies, we extracted again signs from the dataset LSE_eSaude_UVIGO from the 2022 Sign Spotting Challenge at ECCV'22 [22]. In particular, to get reasonable results, we extracted two datasets with two and three signers, respectively, who covered a reasonable number of signs for each class: **ASSDataset1**, with three signers, 10 classes, and at least 9 samples per signer for each class (652 signs in total), and **ASSDataset2**, with two signers, 16 classes, and at least 8 samples per signer for each class (652 signs in total). The protocol is identical to that employed for ACS experiments: results are shown in figure 3, following the



(a)

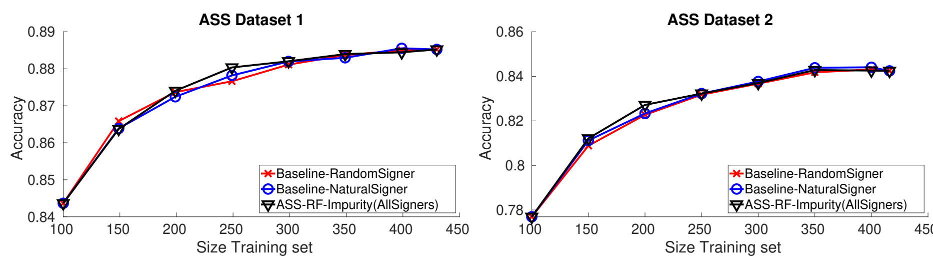


Fig. 3. Active Signer Selection results on the two datasets: (a) comparison between the different ACS strategies; (b) comparison between the best ASS strategy and conventional methods.

same formatting of figure 2: in part (a) we reported the comparison between the different ASS methods, for the two datasets. It seems that the four strategies are almost equivalent: this is confirmed by the statistical tests, from which it holds that, on dataset 1, the different variants are all equivalent according to a t-test followed by a Bonferroni correction; the same holds for dataset 2, except for ASS-RF-Impurity(Allsigners) which is better than alternatives in the central range. In part (b) we reported the comparison between ASS-RF-Impurity(Allsigners) and the standard baselines Baseline-RandomSigner and Baseline-NaturalSigner. In this case the advantage is less evident, with the ASS strategy not permitting to improve conventional strategies: the only statistically significant improvement is in the central part of the plot (250 and 200 objects in the training set for dataset 1 and 2, respectively). Probably, a strategy based on class information is not enough, and we should derive methods which more explicitly exploit signer information – this being the object of current research.

6 Conclusions

In this paper we investigated the usefulness, in the Sign Language Recognition scenario, of ideas and tools of Active Class Selection for the construction of a dataset. We investigated both standard and novel ACS approaches, with experiments on a recent, challenging dataset. Results are promising, showing that a proper dataset construction strategy can be very beneficial for the accuracy of automatic SLR systems. Moreover, we also preliminary investigate other possible exploitations of ACS ideas in this scenario, for example to select which would be, for the classification system, the most beneficial *signer*.

Acknowledgements This research has been funded by the Spanish Ministry of Science and Innovation through the project PID2021-123988OB-C32 and by the Galician Regional Government under project ED431B 2021/24GPC. Manuel Vazquez is also funded by the Spanish Ministry of Science and Innovation through the predoc grant PRE2019-088146. This work has been started during the visit of Manuele Bicego to the University of Vigo in 2022, partially supported by Fondazione Cariverona under the NEUROCONNECT grant.

References

1. Agarwal, A., Thakur, M.K.: Sign language recognition using microsoft kinect. In: Proc. Int. Conf. on Contemporary Computing. pp. 181–185 (2013)
2. Albanie, S., Varol, G., Momeni, L., Afouras, T., Chung, J.S., Fox, N., Zisserman, A.: Bsl-1k: Scaling up co-articulated sign language recognition using mouthing cues. In: Proc. ECCV. pp. 35–53 (2020)
3. Breiman, L.: Random forests. Machine Learning **45**, 5–32 (2001)
4. Bunse, M., Morik, K.: Certification of model robustness in active class selection. In: Proc. ECML-PKDD. pp. 266–281 (2021)
5. Bunse, M., Weichert, D., Kister, A., Morik, K.: Optimal probabilistic classification in active class selection. In: Proc. ICDM. pp. 942–947 (2020)
6. Cakmak, M., Thomaz, A.L.: Designing robot learners that ask good questions. In: Proc. Int. Conf. on Human-Robot Interaction. pp. 17–24 (2012)
7. Chuan, C.H., Regina, E., Guardino, C.: American sign language recognition using leap motion sensor. In: Proc. Int. Conf. on Machine Learning and Applications. pp. 541–544 (2014)
8. Cooper, H., Holt, B., Bowden, R.: Sign language recognition. Visual Analysis of Humans: Looking at People pp. 539–562 (2011)
9. Dietterich, T.G., Lathrop, R.H., Lozano-Pérez, T.: Solving the multiple instance problem with axis-parallel rectangles. Artificial intelligence **89**(1-2), 31–71 (1997)

10. Docío-Fernández, L., Alba-Castro, J.L., Torres-Guijarro, S., Rodríguez-Banga, E., Rey-Area, M., Pérez-Pérez, A., Rico-Alonso, S., García-Mateo, C.: LSE_UVIGO: A multi-source database for Spanish Sign Language recognition. In: Proc. Int. Workshop on the Representation and Processing of Sign Languages. pp. 45–52 (2020)
11. Grishchenko, I., Bazarevsky, V., Zanfir, A., Bazavan, E.G., Zanfir, M., Yee, R., Raveendran, K., Zhdanovich, M., Grundmann, M., Sminchisescu, C.: Blazepose GHUM holistic: Real-time 3d human landmarks and pose estimation. CoRR **abs/2206.11678** (2022)
12. Jedlička, P., Krňoul, Z., Kanis, J., Železný, M.: Sign language motion capture dataset for data-driven synthesis. In: Proc. Int. Conf. on Language Resources and Evaluation (2020)
13. Kottke, D., Krempl, G., Stecklina, M., von Rekowski, C.S., Sabsch, T., Minh, T.P., Deliano, M., Spiliopoulou, M., Sick, B.: Probabilistic active learning for active class selection. In: NeurIPS Workshop on the Future of Interactive Learn. Mach. (2016)
14. Liu, S., Ding, W., Gao, F., Stepinski, T.F.: Adaptive selective learning for automatic identification of sub-kilometer craters. Neurocomputing **92**, 78–87 (2012)
15. Lomasky, R., Brodley, C.E., Aernecke, M., Walt, D., Friedl, M.: Active class selection. In: Proc. of ECML. pp. 640–647 (2007)
16. Naik, N., Madani, A., Esteva, A., Keskar, N.S., Press, M.F., Ruderman, D., Agus, D.B., Socher, R.: Deep learning-enabled breast cancer hormonal receptor status determination from base-level h&e stains. Nature communications **11**(1), 1–8 (2020)
17. Parsons, T.D., Reinebold, J.L.: Adaptive virtual environments for neuropsychological assessment in serious games. IEEE Tr. on Consumer Electronics **58**(2), 197–204 (2012)
18. Rastgoo, R., Kiani, K., Escalera, S.: Sign Language Recognition: A Deep Survey. Expert Systems with Applications **164**, 113794 (2021)
19. Rodriguez-Lujan, I., Fonollosa, J., Vergara, A., Homer, M., Huerta, R.: On the calibration of sensor arrays for pattern recognition using the minimal number of experiments. Chemometrics and Intelligent Laboratory Systems **130**, 123–134 (2014)
20. Settles, B.: Active learning literature survey. Tech. rep., University of Wisconsin-Madison Department of Computer Sciences (2009)
21. Vazquez-Enriquez, M., Alba-Castro, J.L., Docío-Fernández, L., Rodriguez-Banga, E.: Isolated sign language recognition with multi-scale spatial-temporal graph convolutional networks. In: CVPR Workshops. pp. 3462–3471 (2021)
22. Vázquez Enríquez, M., Castro, J.L.A., Fernandez, L.D., Jacques Junior, J.C.S., Escalera, S.: Eccv 2022 sign spotting challenge: Dataset, design andÂăresults. In: ECCV 2022 Workshops. pp. 225–242 (2023)
23. Vazquez-Enriquez, M., Losada-Rodríguez, P., González-Cid, M., Alba-Castro, J.L.: Deep learning and collaborative training for reducing communication barriers with deaf people. In: Proc. Conf. on Information Technology for Social Good (2021)
24. Wu, D., Lance, B.J., Parsons, T.D.: Collaborative filtering for brain-computer interaction using transfer learning and active class selection. PloS one **8**(2), e56624 (2013)
25. Wu, D., Parsons, T.D.: Active class selection for arousal classification. In: Proc. Int. Conf on Affective Computing and Intelligent Interaction. pp. 132–141 (2011)
26. Wu, J., Tian, Z., Sun, L., Estevez, L., Jafari, R.: Real-time american sign language recognition using wrist-worn motion and surface emg sensors. In: Proc. Int. Conf. on Wearable and Implantable Body Sensor Networks. pp. 1–6 (2015)
27. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: Proc. AAAI (2018)