Weak-Supervision Based on Label Proportions for Earth Observation Applications from Optical and Hyperspectral Imagery

Laura E. Cué La Rosa^{1,2}, Dário A. Borges Oliveira^{3,4}, Sam Thiele², Pedram Ghamisi^{2,5} and Richard Gloaguen²

¹Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil

²Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute Freiberg for Resource Technology, Freiberg, Germany

³Data Science in Earth Observation, Technical University of Munich (TUM), Munich, Germany

⁴School of Applied Mathematics, Getulio Vargas Foundation, Rio de Janeiro, Brazil

⁵Institute of Advanced Research in Artificial Intelligence (IARAI), 1030 Vienna, Austria

Abstract

In this paper, we assess a weak-supervised approach that employs weak constraints in the form of class proportions to train a neural network capable of performing pixel-wise classification for Earth Observation (EO) applications. The approach combines self-supervised contrastive clustering and a constraint on cluster proportions in an online fashion allowing its application in large-scale EO images. The methodology is based on the generation of simple augmented views of input image tiles, and the use of a loss function that performs contrastive learning to achieve consistent results that are invariant to these augmentations, and simultaneously follow the cluster proportions constraint. In many EO applications, information about class proportions is available through expert knowledge or e.g., governmental census. This weak information about class proportions allows training a classifier without information about the class at the pixel-level, alleviating the burden of manual annotation. In this context, crop and geological mapping from EO data are two crucial applications in the search for sustainable ways of resource management. We tested the approach upon optical and hyperspectral data achieving promising results and proving the method's applicability across different applications and data sources.

Keywords

Weak-supervision, Learning from proportions, Multi-source, Crop mapping, Geological mapping.

1. Introduction

Self-supervised learning [1, 2, 3, 4] has recently emerged as a powerful tool in computer vision applications. Among the existing self-supervised methods, contrastive learning can be considered the most promising one. This type of approach is based on the generation of augmented versions of the input image and the use of a twin network that performs feature extraction that combined with a loss function performs contrastive learning to achieve consistent results between these augmentations. The contrastive loss function is expected to increase the similarity among the augmentations of the same image while decreasing the similarity from augmentations of different images. The

lauracuerosa@gmail.com (L. E. C. L. Rosa);

darioaugusto@gmail.com (Dário A. B. Oliveira); sam.thiele01@gmail.com (S. Thiele); p.ghamisi@gmail.com

(P. Ghamisi); r.gloaguen@hzdr.de (R. Gloaguen)

(R. Gloaguen) © Commons License Attribution 4.0 International (CC BY 4.0). CEUR Workshop Proceedings (CEUR-WS.org) main characteristic of these methods is the capability of learning meaningful feature representations in an unsupervised fashion. This capability has opened new venues in other research fields beyond computer vision such as Earth Observation (EO) applications. In this context, crop and geological mapping from EO data are two crucial applications to agricultural monitoring and modern mining, where frequently limited or non-existent training information is available.

Considering EO applications, self-supervised methods have been employed with success including image classification, object detection and semantic segmentation [5, 6, 7, 8, 9]. Some of these works employ geolocation and spatio-temporal information to learn a more discriminative set of features for remote sensing applications [5, 10]. Hyperspectral image classification and clustering using contrastive learning have also been the focus of recent publications [9, 8]. However, all the approaches mentioned above need positive and negative sample pairs to perform the contrastive loss, which is computationally intensive.

One of the most important contrastive-learning methods is the *Swapping Assignments between Multiple Views* (SwAV) [2], which performs self-supervised and clustering in an online fashion. The method employs an

CDCEO 2022: 2nd Workshop on Complex Data Challenges in Earth Observation, July 25, 2022, Vienna, Austria

 ^{0000-0002-6284-9494 (}L. E. C. L. Rosa); 0000-0002-0674-5332 (Dário A. B. Oliveira); 0000-0003-4169-0207 (S. Thiele); 0000-0003-1203-741X (P. Ghamisi); 0000-0002-4383-473X (R. Cloamen)

optimal transport (OT) solver to assign the image feature vectors to cluster centroids by means of an equipartition constraint that ensures that all samples within a batch of images are equally assigned to the predefined number of clusters.

An advantage of the SwAV method over the previously proposed contrastive learning frameworks is that the use of the OT solver with the equipartition constraint allows disregarding pairwise comparisons. Recently, weak information in the form of class proportions was introduced as a constraint in SwAV to train a classifier in a weakly-supervised fashion. The method called *Learning from Label Proportions with Prototypical Contrastive Clustering* (LLP-Co) [11] disregards the equipartition constraint in the OT solver by adding a cluster proportions constraint.

Using information about class proportions to train a classifier has gained more attention in the last years [12, 13, 14, 15]. Given a set of images, Learning from Label Proportions (LLPs) approach focuses on learning an instance-level classifier using as reference signal only the class proportions observed in this set. In EO applications, with a large amount of available data and the unavailability of pixel-level annotations, the use of priors like class proportions is an attractive solution. In many real-life scenarios, these proportions can be obtained by governmental census or even expert knowledge. Examples of governmental agencies that record statistics about agriculture, forestry, and natural resources, among others, are the National Agricultural Statistics Service of the United States Department of Agriculture¹, the Brazilian Institute of Geography and Statistics (IBGE) in Brazil², Forest Research in the United Kingdom³, and the European Statistics website ⁴.

This paper focuses on accessing the viability of using contrastive learning combined with LLP to train a pixelwise classifier based only on prior information about global class proportions for EO applications. We tested the LLP-Co methodology upon two datasets, the first focuses on crop type mapping using optical data and the second on geological mapping using hyperspectral data. This allows assessing the model's applicability across different applications and data sources. Hence, the main contribution of this study is to propose a weaksupervised deep clustering method that employs label proportions as priors and can be easily applied to largescale EO data from different sources for significantly different applications.

¹https://www.nass.usda.gov/

²https://www.ibge.gov.br/

 $^{3} https://www.forestresearch.gov.uk/tools-and-resources/statistics/forestry-statistics/$

⁴https://ec.europa.eu/eurostat

2. METHOD

2.1. LLP and Optimal Transport

In this work, we asses the LLP-Co approach in a scenario where only to the global class proportions are available to train the network. To implement LLP, the training samples are split into S disjoint bags of image tiles, where B_i is the *i*th bag, which consists of a set of s_i randomly cropped image tiles from the large scale input EO image. Here, $\mathcal{B}_i = \{(\mathbf{x}_{i,j})\}_{i=1}^{s_i}$, where $\mathbf{x}_{i,j}$ is the image tile j within the bag i. The final training set is then expressed as $\mathcal{T} = \{(\mathcal{B}_i, \mathbf{w})\}_{i=1}^S$, where **w** is a vector of global label proportions, which is the same for all bags B_i . In a multi-class problem with K classes, $\mathbf{w} \in \Delta_K$ and s.t. $\sum_{k=1}^{K} \mathbf{w}^{k} = 1$, where the \mathbf{w}^{k} element is the proportion of tiles that belong to class k. In the methodology a neural network acts as the feature extractor followed by layer that delivers the class probabilities vector $\tilde{\mathbf{p}}_{i,j} =$ $p_{\theta}(\mathbf{y}|\mathbf{x}_{i,j})$, where θ represents the network parameters [16]. Then, the estimated global label proportions for each bag is expressed as:

$$\label{eq:window} \hat{\mathbf{w}}_i = \frac{1}{s_i} \sum_{j=1}^{s_i} \tilde{\mathbf{p}}_{i,j},$$

and to train the network a standard cross-entropy loss function can be used

$$L(\hat{w}, w) = -\frac{1}{S} \sum_{i=1}^{S} \mathbf{w} \log \hat{\mathbf{w}}_{i}.$$
 (1)

The above equation is reformulated by encoding the label proportions as a posterior distribution [1, 17, 11]

$$L(p,q) = -\frac{1}{S} \sum_{i=1}^{S} \sum_{j=1}^{s_i} \sum_{k=1}^{K} \frac{q(y^k | \mathbf{x}_{i,j})}{s_i} \log p_{\theta}(y^k | \mathbf{x}_{i,j})$$
(2)

delivering the LLP optimization objective as:

$$\min_{(p,q)} L(q,p), \quad \text{s.t.} \quad \forall y : q(y^k|\cdot) \in [0,1]$$
(3)

$$\sum_{j=1}^{s_i} q(y^k | \mathbf{x}_{i,j}) = \mathbf{w}^k s_i, \tag{4}$$

where the global proportion constraint ensures that each label k contains overall $\mathbf{w}^k s_i$ samples. This equation is an instance of the regularized optimal transport problem and is solved using the Sinkhorn-Knopp algorithm [1, 17, 11]. Here $\mathbf{P}_{i,j}^y = p_\theta(y|\mathbf{x}_{i,j}) \frac{1}{n_i}$ is the probabilities matrix estimated by the network and $\mathbf{Q}_{i,j}^y = q(y|\mathbf{x}_{i,j}) \frac{1}{n_i}$ is the matrix of assigned probabilities for bag \mathcal{B}_i . In the LLP-Co approach, \mathbf{Q}_i splits the samples within the bag following the global label proportions. Then the objective function as an OT solver is defined as

$$\min_{\mathbf{Q}_i \in U(\mathbf{w}, \mathbf{a}_i)} \langle \mathbf{Q}_i, -\log \mathbf{P}_i \rangle + \varepsilon h(\mathbf{Q}_i), \qquad (5)$$

where $U(\mathbf{w}, \mathbf{a}_i)$ is the matrix space of possible solutions for the *i*-th bag, and $\mathbf{a} = (1/n_i)\mathbf{1}_{n_i}$ is a normalizing constraint [18].

2.2. Learning from Global Label Proportions with Prototypical Contrastive Clustering

LLP-Co [11] is a self-supervised contrastive method that performs online clustering by means of a convolutional neural network that delivers consistent cluster assignments between augmentations of the same input. At the same time, the cluster assignment must follow certain cluster size constraints that are provided as weak information. Given a user-defined number of views of the same input image tile, the algorithm employs the OT solver in Eq.5 to compute soft targets or codes. These targets as then considered as true labels to calculate the cross-entropy considering the network's prediction for The methodology pipeline for two other views. augmented views and K classes is the following. First each image tile j within a bag is transformed into two augmented version fed to an encoder network that extracts the features vectors $\mathbf{z}_{i,j}^{t1}, \mathbf{z}_{i,j}^{t2}$. These features are then mapped to one of K trainable prototypes ${\bf V}$ to perform the code assignments for each view $\mathbf{c}_{i,j}^{t1}$ and $\mathbf{c}_{i,j}^{t_2}$ using the OT solver. From then on, a "swapped" contrastive loss is applied to predict the assignment of one feature from the code of the other. The optimization process is then conducted by minimizing the loss for all samples j within bag i:

$$\mathcal{L}_{swap}(\mathbf{z}_{i,j}^{t1}, \mathbf{z}_{i,j}^{t2}) = \ell(\mathbf{z}_{i,j}^{t1}, \mathbf{c}_{i,j}^{t2}) + \ell(\mathbf{z}_{i,j}^{t2}, \mathbf{c}_{i,j}^{t1}), \quad (6)$$

where each term is the cross-entropy loss between the code and the probability obtained after applying a softmax function on the dot product between the features \mathbf{Z}_i and the prototypes **V**. For more information about the LLP-Co method, see [11].

3. Datasets

1

3.1. Campo Verde dataset (CV)

The first study site is in Campo Verde, an agricultural region located in Mato Grosso, at a latitude of $15^{\circ}32'48''$ south and a longitude of $55^{\circ}10'08''$ west, Brazil (Fig. 1). Campo Verde (CV) [19] is a public dataset ⁵ that provides pre-processed SAR and Optical images between October 2015 and July 2016. The major crops found in the region are *soybean*, *maize* and *cotton*. Other crops and non crops categries are *beans*, *sorghum*,

Non-Commercial Crops (NCC), *pasture, eucalyptus, turfgrass, cerrado* and *soil.* This work focuses in the second seeding period for major crops *maize* and *cotton* for months between March to July. The reference data consisted of 608 parcels. Table 1 gives the percentages of the overall area planted with major crops accordingly to the annotated parcel, we use this information as the global vector of class proportions for our experiments.



Figure 1: Overview map of Brazil, Mato Grosso state, and the Compo Verde region were the images were acquired.

3.2. Corta Atalaya dataset (CA)

The second study area is located at Rio Tinto, Spain. Rio Tinto is located 70 km north of Huelva in the Iberian Pyrite Belt (IPB), a belt extending from southern Portugal into southern Spain (Fig. 2). Our data was collected from Corta Atalaya (CA), an open-pit mine with a size of 1200 \times 900 m and a depth of ca. 350 m. This pit exposes basaltic to intermediate volcanic rocks along the northern part of the pit, and overlying felsic volcanic rocks, slate, and conglomerate which are exposed in the western part of the mine. We tested our approach using ground-based hyperspectral imagery collected using a tripod-mounted Specim AsiaFENIX sensor, which covers the visible-near and short-wave infrared range. A labeled reference image was created based on field mapping, fifty-seven hand samples, and combined supervised classification followed by manual interpretation of the hyperspectral data [20]. The lithologies interpreted at CA are as follows: oxidised, massive sulphide, two varieties of chlorite, two sericitic units, shale and purple shale. In this study, we grouped the lithologies into two major categories, chlorite schist and mineralised volcanics, in addition, weathered material and vegetation were grouped in a category named others. Table 1 gives the percentages of the overall area with these two major lithologies accordingly to the labeled reference image, we use this information as the global vector of class proportions for our experiments. For more

⁵The CV database is available from IEEE Dataport at https: //ieee-dataport.org/documents/campo-verde-database.

information about the dataset, we refer the reader to [20].



Figure 2: Overview map of the Iberian Pyrite belt (a) with locations of the main volcano-sedimentary units (green). The geology of the Corta Atalaya and Cerro Colorado open pits is also shown (b). Maps taken with permission from [20].

Table 1

Global class proportions (%) for each dataset accordingly to the reference data. Cs standsd for *chlorite schist* and Mv stands for *mineralised volcanics*.

	CV			СА		
Cotton	Maize	Others	Cs	Mv	Others	
45.3	35.8	18.9	38.7	57.7	3.6	

4. Experiments

4.1. Experimental Protocol

Our experiments focused on the major categories found in both datasets. To assess the methodology's robustness to different data sources, we employed optical data for CV dataset and hyperspectral data for CA dataset. For the CV dataset, we considered the cloud-free optical image available for May 2016. For the CA dataset, we stacked VNIR and SWIR data in a unique data cube. We evaluated the LLP-Co method under a scenario that uses global class proportions to identify the major categories in the target regions. Unlike the traditional LLP training schemes, which calculate the class proportion for each bag of samples independently in a supervised way, our proposal uses only weak information.

In our experiments, we used as prior information the global proportions reported in Table 1. Given the bag size s_i , we defined the training bag \mathcal{B}_i by randomly cropping s_i image tiles from the large-scale images. The tiles were cropped from the annotated area and we used the class of the central pixel of the tile. As the bag size increases, the class proportions within the bag converge to the global class proportions found in the dataset, hence we adopted a large bag size of $n_i = 2048$ for both datasets.

4.2. Implementation Details

Considering the different data sources, we employed a modified ResNet18 and ResNet10 as the backbone architecture for CV and CA datasets, respectively. To process the hyperspectral data cube in both spatial and spectral domains with also added two 3D convolutional layers at the beginning of the ResNet10 network for the CA dataset. The ResNet architecture is then followed by a projection head that projects the features to a 1024-dimensional space. We trained the models for 100epochs using stochastic gradient descent with cosine learning rate decay [21]. The image tiles size was set to 21×21 for both datasets. For each dataset, we randomly selected 200,000 image tiles on the fly to create the random bags. The list of augmentations includes random rotations, mirroring, and random resizing to obtain two views. For the OT solver, we set the hyper-parameters as in [11]. The number of clusters for both models was set to the number of categories found in the datasets. We quantitatively assessed the method using three metrics: cluster accuracy (Acc), macro average F1-score (F1-score), and normalized mutual information (NMI). Since we use the class proportion information, we reported the classification metrics by considering the cluster assigned by the network at inference time. We also report the confusion matrices.

4.3. Baseline method

We adopted the original SwAV method with the equipartition constraint as the baseline method. This constraint ensure that samples are equally partitioned among the clusters, and for a good performance the authors recommend a number of cluster at least three times higher than the expected number of categories. In preliminary experiment we found that 30 cluster delivered a good performance for CV dataset, while 10 cluster delivered an acceptable performance for CA dataset. The backbone network for SwAV is the same as the LLP-Co backbone network for each dataset. To evaluate the model we used the feature z generated by the backbone network followed by a k-means clustering.



Figure 3: Maps of the class output CV and CA datasets. Crop types for CV dataset: *maize*, *cotton*, *cott*

An Hungarian match [22] between the true categories and the k-means result delivered the final accuracy.

Fabl	e 2					
Fest	performance	for the	CV and	CA	datasets	

5. Results

Table 2 shows the performance for both datasets in terms of Acc, F1-score, and NMI. The model performance reported competitive results, achieving accuracies of 94.1% and 91.6% for the CV and CA datasets, respectively. Similar performance was observed in terms of F1-score for CV dataset with 93.8%. In contrast, for CA dataset, a lower value was observed with 76.9% of F1-score due principally to class others. The cluster quality metrics NMI reported values of 0.76 and 0.66 for CV and CA, respectively. Considering these metrics, the CV dataset reported better results than CA dataset. This may be due to the different types of application and data since geological mapping from hyperspectral data is a more challenging task due to significant confounding data variance and often subtle distinctions between the features of interest.

Comparing LLP-Co with the baseline model, we observe that, as expected, the inclusion of priors into the training process was crucial for a good classification performance. LLP-Co outperformed SwAV by ~20% and ~30% in terms of accuracy for the CV and CA datasets, respectively. Similar improvement was observed for the F1-score, achieving an enhancement of ~27% and ~30% for CV and CA datasets, respectively.

Table 3 presents the confusion matrices. As expected, the per-class accuracy achieved high performance for

Metric	LLP-Co		SwAV	
, notifie	CV	CA	CV	CA
Acc	94.1%	91.6%	74.4%	61.0%
F1-score	93.8%	76.9%	66.0%	47.5%
NMI	0.76	0.66	0.50%	0.38

the major categories, with values above 91% for both datasets. However, in CA dataset, 48% of class *others* was misclassified as *chlorite schist*, demonstrating the challenge of this task. Another possible explanation of this drop in performance can be related to the distribution of the classes, since considering a more balanced vector of class proportions (like in CV dataset with $\mathbf{w} = (45.3, 35.8, 18.9)$) but significantly different among the classes, delivers much better performance, allowing the model to learn a more discriminative and relevant set of features. In contrast, for a highly unbalanced vector of proportions, the model will favor the majority classes, as we observed for the CA dataset.

Finally, Fig. 3 presents the classification maps for each dataset. Here we can observe classification errors between class *maize* and the other two classes for CV dataset, and class *mineralised volcanics* with class *others* for CA dataset. In addition, it is worth pointing out the quality of the predictions for both datasets, where no salt-and-pepper effect was observed.

 Table 3

 LLP-Co confusion matrices for the CV and CA datasets for major categories and class others.

			Predicted	
	CV	Maize	Cotton	Others
	Maize	91%	5%	4%
Lu	Cotton	2%	96%	2%
	Others	2%	3%	95%
			Predicted	
	CA	Cs	Predicted Mv	Others
e	CA Cs	Cs 94%	Predicted Mv 5%	Others 1%
[rue	CA Cs Mv	Cs 94% 2%	Predicted Mv 5% 93%	Others 1% 5%

6. Conclusions

This recently work evaluates а proposed weak-supervised method that combines contrastive learning with class proportions constraints to train a classifier without the need for labels at the pixel level in the context of Earth Observation (EO) applications. The approach was able to archive reasonable accuracy values across different tasks and data sources, proving its robustness and applicability to large-scale EO data. Overall accuracy of 90% was reported for crop and geological mapping applications considering the major categories found in the target regions. The approach also failed to identify classes with very small proportions. Several ways of dealing with this problem such as weighted cross-entropy or focal loss can be also implemented into our method. The success of the methodology opens a new path in the use of weak information to help alleviate the burden of manual annotation in EO.

References

- Y. M. Asano, C. Rupprecht, A. Vedaldi, Self-labelling via simultaneous clustering and representation learning, arXiv preprint, arXiv:1911.05371 (2019).
- [2] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, A. Joulin, Unsupervised learning of visual features by contrasting cluster assignments, Advances in Neural Information Processing Systems 33 (2020) 9912–9924.
- [3] J. Li, P. Zhou, C. Xiong, R. Socher, S. C. Hoi, Prototypical contrastive learning of unsupervised representations, arXiv preprint, arXiv:2005.04966 (2020).
- [4] C. Li, X. Li, L. Zhang, B. Peng, M. Zhou, J. Gao, Self-supervised pre-training with hard examples improves visual representations, arXiv preprint, arXiv:2012.13493 (2020).

- [5] K. Ayush, B. Uzkent, C. Meng, K. Tanmay, M. Burke, D. Lobell, S. Ermon, Geography-aware selfsupervised learning, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10181–10190.
- [6] W. Li, H. Chen, Z. Shi, Semantic segmentation of remote sensing images with self-supervised multitask representation learning, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14 (2021) 6438–6450.
- [7] V. Stojnic, V. Risojevic, Self-supervised learning of remote sensing scene representations using contrastive multiview coding, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 1182–1191.
- [8] Y. Cai, Z. Zhang, Y. Liu, P. Ghamisi, K. Li, X. Liu, Z. Cai, Large-scale hyperspectral image clustering using contrastive learning, arXiv preprint, arXiv:2111.07945 (2021).
- [9] J. Yue, L. Fang, H. Rahmani, P. Ghamisi, Selfsupervised learning with adaptive distillation for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 60 (2021) 1–13.
- [10] O. Mañas, A. Lacoste, X. Giro-i Nieto, D. Vazquez, P. Rodriguez, Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 9414– 9423.
- [11] L. E. C. L. Rosa, D. A. B. Oliveira, Learning from label proportions with prototypical contrastive learning, in: to appear, AAAI, 2022.
- [12] Z. Qi, B. Wang, F. Meng, L. Niu, Learning with label proportions via NPSVM, IEEE Transactions on Cybernetics 47 (2016) 3293–3305.
- [13] G. Dulac-Arnold, N. Zeghidour, M. Cuturi, L. Beyer, J.-P. Vert, Deep multi-class learning from label proportions, arXiv preprint, arXiv:1905.12909 (2019).
- [14] Y. Shi, J. Liu, B. Wang, Z. Qi, Y. Tian, Deep learning from label proportions with labeled samples, Neural Networks 128 (2020) 73–81.
- [15] C. Scott, J. Zhang, Learning from label proportions: A mutual contamination framework, Advances in Neural Information Processing Systems 33 (2020) 22256–22267.
- [16] J. Liu, B. Wang, Z. Qi, Y. Tian, Y. Shi, Learning from label proportions with generative adversarial networks, Advances in Neural Information Processing Systems 32 (2019) 7169–7179.
- [17] J. Liu, B. Wang, X. Shen, Z. Qi, Y. Tian, Two-stage training for learning from label proportions, arXiv preprint, arXiv:2105.10635 (2021).
- [18] A. Genevay, G. Dulac-Arnold, J.-P. Vert,

Differentiable deep clustering with cluster size constraints, arXiv preprint, arXiv:1910.09036 (2019).

- [19] I. D. Sanches, R. Q. Feitosa, P. M. A. Diaz, M. D. Soares, A. J. B. Luiz, B. Schultz, L. E. P. Maurano, Campo Verde database: Seeking to improve agricultural remote sensing of tropical areas, IEEE Geoscience and Remote Sensing Letters 15 (2018) 369–373.
- [20] S. T. Thiele, S. Lorenz, M. Kirsch, I. C. C. Acosta, L. Tusa, E. Herrmann, R. Möckel, R. Gloaguen, Multi-scale, multi-sensor data integration for automated 3-d geological mapping, Ore Geology Reviews 136 (2021) 104252.
- [21] I. Loshchilov, F. Hutter, Sgdr: Stochastic gradient descent with warm restarts, arXiv preprint, arXiv:1608.03983 (2016).
- [22] H. W. Kuhn, The Hungarian method for the assignment problem, Naval Research Logistics Quarterly 2 (1955) 83–97.