SELF-SUPERVISED PRETRAINING FROM HANDCRAFTED FEATURES FOR CHEST X-RAY CLASSIFICATION

K. Yematsinau[1], V. Kovalev[2]

[1] Belarusian State University, Minsk, Belarus
[2] United Institute of Informatics Problems of NAS, Minsk, Belarus
E-mail: {emationov.key, vassili.kovalev}@gmail.com

Modern convolutional neural networks require a large amount of human labeled data during training process. Prior work demonstrates that this problem can be addressed using self-supervised learning. This paper presents a novel self-supervised pretraining approach, which has been shown to be beneficial for the quality and stability of training process in case of domain-specific datasets with a small amount of labeled data.

Keywords*: neural networks, biomedical images, self-supervised learning.*

Introduction. The effectiveness of machine learning methods generally depends on the properly prepared feature representations of the original data. In particular, the transformation of the input images in order to obtain representative features for the specific algorithm was one of the most important stages throughout the development of the fields of computer vision and biomedical image analysis [1].

In recent decades, methods based on the use of convolutional neural networks (CNN), for which applying directly to the image is possible, have become more widespread. However, they are not without drawbacks: in order to achieve high quality, classical CNN require a large amount of human labeled data [2]. In practice, for many tasks in computer vision, particularly in medical images analysis domain, there is a rather limited number of labeled images and almost unlimited number of unlabeled ones. In this regard, it is relevant to study methods for learning informative features without any external labels. This group of methods is often called self-supervised or unsupervised representation learning.

In this work, we propose a novel approach to pretrain CNNs in self-supervised manner, which uses years of human experience in a novel way: without explicit use of expert-assigned labels. The method can be used in order to improve the stability of training in case of datasets with a small amount of labeled data.

Preliminaries. Pretraining is the process that utilizes large general purposed datasets be to improve performance on downstream datasets/tasks. Common type of pretraining is strongly associated with transfer learning approach: model weights trained on source data and used to initialize training process on the target dataset/task. Model performance generally scales with source dataset size and the similarity between the source and target da-

tasets/tasks. A fundamental challenge for transfer learning is to improve the performance on target data when it is not similar to source data [3].

Self-supervised pretraining is a form of unsupervised training that captures the informative features and patterns without human-provided labels. Most often researches extracts some labels, that are naturally aligned to the source dataset, or transform source data and map labels in accordance to the transformations applied [4, 5]. Then the model is trained to predict these "automatically aligned" labels. Such task is often called pretext task and generated labels – pseudo labels.

As a consequence, in contrast to usual "general-purpose" dataset pretraining self-supervised methods are naturally not prone to the dissimilarity of source/target data: in described setup source and target dataset are based on the same images, but target and pre-text tasks have different labels.

Classical handcrafted features have been investigated for decades and, as a result, corresponding extraction algorithms contains a vast amount of accumulated human experience: researchers have worked over the years in order to learn how to extract informative and relevant features for the specific tasks. Some examples of such features are LBP histograms, Haralick texture features, aggregated co-occurrence matrices [6]. In our work we use such methods along with an extensive experience they are providing to automatically extract supervision signal from images for pretraining of modern deep learning models.

Method. Inspired by recent self-supervised methods, based on pseudo labels generation and assignment, our approach also relies on the classical feature extraction algorithms. We propose to use such features as pseudo labels for the in domain pretraining of the convolutional neural network. An overview of the investigated pretraining approach is presented in Fig 1.

1. Extract handcrafted features (e.g. LBP histograms) from the original images

Generate pseudo labels based on extracted features. To prepare multi-label classification pretext task's labels they can be quantized or clustering algorithms can be used.

Formulate the pretext task based on generated labels and train target convolutional neural network in a standard manner (minimizing error for pretext classification task).

The intuition behind self-supervised techniques is that trained model need to learn significant information from images to solve corresponding pretext tasks. Considered that during preparation of the pretext task for our approach labels are based on meaningful handcrafted features, model supposed to learn semantic-related information to correctly classify such labels.

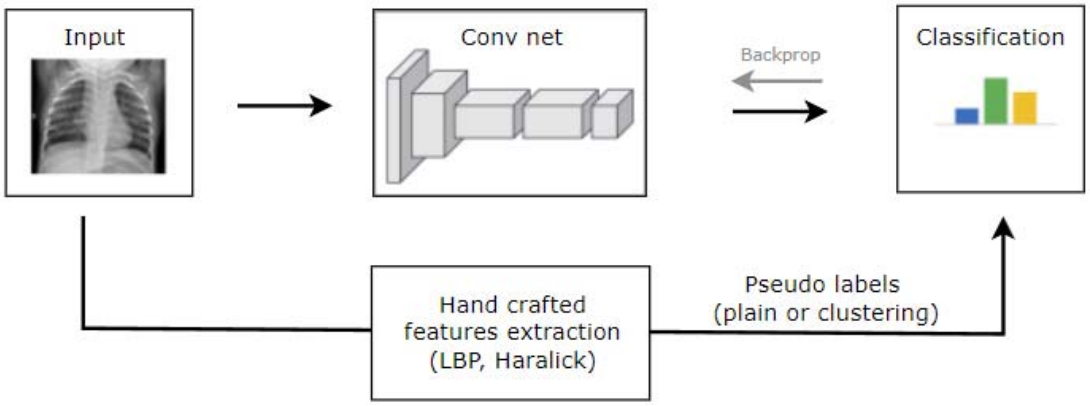

Fig 1. Illustration of the proposed pretraining method: we extract handcrafted features and use them as pseudo-labels for convolutional neural network pre-training process

For the sake of simplicity and interpretability in this paper we use clustering algorithms (K-MEANS, DBSCAN [7]) to generate simple and meaningful labels from extracted handcrafted for proposed pretext task. Investigation of other approaches (e.g. quantization) requires future research.

Experiments. To validate proposed method, we set up the experiment on Cell Chest X-Ray dataset, which comprises 5,323 X-ray images from children, including 3,883 cases of viral and bacterial pneumonia and 1,349 normal images [8].

First, we generate labels for the proposed self-supervised task: extracted Haralick texture features from dataset's images and clustered them using DBSCAN algorithm. As a result, we got 5 different cluster labels for pretext task. Then we pretrained ResNet-50 model on proposed pre-text task. The effective batch size was set to 64. We used the Adam optimizer with a learning rate of 0.0001. The model was trained for 20 epochs without stopping criteria as described in [9]. Using the same parameters we have also pretrained additional model, which was initialized by Image-net pretrained weights [10]. For the pretext task 96.8% and 98.1 % accuracy respectively were achieved.

Next, we setup three separate fine-tuning experiments: we trained only last (classification) layer of Image-net pretrained, self-supervised pretrained and combined pretrained models (which we prepared during previous step). The training parameters were the same as at the pretraining step, except 0.001 learning rate.

The metrics are presented in Table 1:

Table 1

Comparison of different pretraining approaches

Image-net Pretrained		Self-supervised Pre-trained (ours)		Combined Pretrained (ours)	
ACC, %	AUC, %	ACC, %	AUC, %	ACC, %	AUC, %
83.2	94.9	83.1	95.8	89.2	96.9

Conclusion. The results show that the proposed method itself can leads to comparable with classical pretraining quality. Higher metrics of combined approach highlight that our technique can be even more successful as a second step after standard pretraining. This outcome proves that such "in domain" pretraining can be highly beneficial and improve quality of the learning process in case of specific datasets (compared to large-scaled general-purpose ones) with a small amount of labeled data.

REFERENCES

1. Bengio Y., Courville A., Vincent P. Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence. 35(8):1798–1828. 2013. DOI:10.1109/TPAMI.2013.50
2. Hestness J., Narang S., Ardalani N. et al. Deep Learning Scaling is Predictable, Empirically [Electronic resource]. – Mode of access: https://arxiv.org/abs/1712.00409. – Date of access: 31.03.2022
3. Zhuang F., Zhiyuan Q., Keyu D. et al. A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE. P. 1-34. 2020. DOI:10.1109/JPROC.2020.3004555.
4. Chen T., Kornblith S., Norouzi M. et al. A Simple Framework for Contrastive Learning of Visual Representations. ICML. 2020. P. 1597–1607.
5. Caron M., Bojanowski P., Joulin A. et al. Deep Clustering for Unsupervised Learning of Visual Features. ECCV. 2018. P. 139-156. DOI: 10.1007/978-3-030-01264-9_9
6. Kovalev V., Petrou M. Multidimensional co-occurrence matrices for object recognition and matching. Graphical models and image processing. CVGIP: Graphical Model and Image Processing. 1996. P. 187-197.
7. Xu D., Tian Y. A Comprehensive Survey of Clustering Algorithms // Ann. Data. Sci. 2, pp. 165–193. 2015. DOI:10.1007/s40745-015-0040-1
8. Kermany D., Zhang K., Goldbaum M. Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification // Mendeley Data. DOI:10.17632/rscbjbr9sj.2
9. Gazda M., Plavka J., Gazda J. et al, Self-Supervised Deep Convolutional Neural Network for Chest X-Ray Classification // IEEE Access. 2021. Vol. 9. P. 151972-151982. DOI: 10.1109/ACCESS.2021.3125324.
10. Deng J., Dong W., Socher R et al. Imagenet: A large-scale hierarchical image database // IEEE conference on computer vision and pattern recognition. 2019. P. 248–255.