REVIEW

of the dissertation
of Xu Jiashu
on the topic «Research and development of self-supervised visual feature learning based on neural networks»,
submitted for earning the degree of Doctor of Philosophy
in the field of knowledge 12 Information Technologies
with the specialty 121 Software Engineering

Relevance of the dissertation topic.

Nowadays, computer vision has become helpful in many fields, such as industry, medicine, security, transport, etc. Classical methods of unsupervised learning face problems in extracting high-quality feature representations. To address the challenge of learning universal feature representations from a large amount of unlabeled data, the approach of self-supervised learning emerged. Although self-supervised learning has made tremendous progress, the development of universal large-scale models in the computer vision field is still in its early stages, and self-supervised learning for image understanding remains challenging. The dissertation research focuses on advancing self-supervised learning methods in the visual domain, specifically for extracting generic visual features without requiring labeled data. By leveraging pre-training models through self-supervised learning, these models can be applied to downstream tasks such as medical image analysis and diagnostics, thereby enhancing their accuracy. Considering the importance of visual data processing, and the wide scope of application, the research is relevant.

Evaluation of the Justification of the Dissertation's Scientific Results, Their Reliability, and Novelty.

The scientific novelty of the dissertation research results lies in the following:

1. A novel self-supervised learning methodology leveraging the Mixup Feature function is introduced. This method involves the pre-training of visual representations by predicting Mixup features from masked images, which stand in as proxies for advanced semantic information. The approach is poised to potentially bolster the aggregate efficacy of the model.

2. A masked autoencoder model for self-supervised learning is proposed, featuring novel mechanisms for noise suppression and self-distillation. The architecture utilizes a masked autoencoder in conjunction with a teacher network to facilitate the reconstruction of corrupted image segments afflicted with random Gaussian noise. This model extends the utility of self-supervised techniques in restoring visual data.

3. For the first time, a model is proposed that, by combining losses at the pixel level and feature level, enables the extraction of deep semantic characteristics of the

image. This complements existing techniques for modeling masked images and also increases the robustness of self-supervised learning models when working with unbalanced data sets.

This thesis primarily addresses the challenge of effectively training deep learning algorithms without reliance on extensively labeled datasets. By exploring self-supervised learning algorithms, it introduces two novel approaches: 'Mixup Feature' and 'Denoising Self-Distillation Masked Autoencoder'. These methods constitute significant additions to the broader family of self-supervised learning algorithms. Notably, they reduce dependence on labeled data while achieving accuracy comparable to supervised visual feature learning methods. In practical applications, particularly in medical image analysis, where acquiring a large volume of labeled medical imagery is often challenging, these algorithms hold immense potential. Given the global shortage of medical experts and the escalating demand for diagnostic services, there is an urgent need for accurately interpreting medical images with minimal human intervention. This work could facilitate the development of cost-effective pre-training schemes that obviate the need for expensive and time-consuming data labeling.

Thus, in the dissertation, the scientific task has been fully accomplished, and the candidate has fully mastered the methodology of scientific activity.

Evaluation of the content of the dissertation, its completeness, and adherence to the principles of academic integrity.

The content of the dissertation of Xu Jiashu fully meets the Higher Education Standard for the specialty 121 Software Engineering and research directions according to the educational program "Software Engineering".

The dissertation is a completed scientific work and demonstrates the presence of the candidate's personal contribution to the scientific direction of Self-supervised Visual Feature Learning.

Having examined the similarity report based on the check of the dissertation for text matches, it can be concluded that the dissertation of Xu Jiashu is the result of the candidate's independent research and does not contain elements of falsification, compilation, fabrication, plagiarism, and borrowing. The ideas, results, and texts of other authors used have proper references.

Language and Style of Presentation of Results.

The dissertation is written in English. It comprises of an introduction, 4 chapters, conclusions, references, and appendices. The total volume of the dissertation is 168 pages.

In the introduction, the significance of Self-Supervised Learning (SSL) and its applications in both natural language processing and computer vision is highlighted. The aim of the research, its key tasks, scientific novelty of the results and practical significance are presented.

In Chapter 1, there is a comprehensive review of the evolution of self-supervised learning algorithms used in visual feature learning. While the methods selected may no longer considered competitive in the contemporary context, these early self-supervision techniques have indelibly shaped the foundation for current strategies. These algorithms are categorized into four main types: contrastive learning, masked image modeling, self-distillation, and canonical correlation analysis. It meticulously examined how these methods collectively contribute to the advancement of self-supervised visual representation learning. This thorough analysis highlights the field's historical progress and sets the stage for understanding current and future trends in self-supervised learning within computer vision.

In Chapter 2, two novel self-supervised learning algorithms are presented. The first algorithm, Mixup Feature, proposes a new pretext task of reconstructing a Mixup of traditional image features like Sobel, HOG, and LBP as the target for a masked autoencoder. This unique mixed feature target provides more complex visual representations for pertaining. The second algorithm, Denoising Self-Distillation Masked Autoencoder, combines self-distillation with masked autoencoder for robust denoising. It considers both pixel-level image restoration and feature-level regression. Gaussian noise is randomly added to image patches as a pretext task for the student network to denoise. The teacher network guides the student through exponential moving average parameter updates. An asymmetric decoder further enhances efficiency. The loss function balances pixel reconstruction loss and feature alignment loss.

In Chapter 3, the results of experiments conducted to verify the proposed Mixup Feature and Denoising Distillation Masked Autoencoder self-supervised learning algorithms are presented. For both algorithms, CIFAR-10, CIFAR-100, and STL-10 datasets are used for the model pre-training. Values of the masking ratio and the mixup factor are experimentally optimized. For these optimal parameters, comprehensive experiments verified the effectiveness of the proposed self-supervised learning algorithms compared to existing methods.

In Chapter 4, the application of self-supervised pre-trained models in the classification of Computerized Tomography scans is considered. Three self-supervised learning methods – MAE, Mixup Feature, and Denoising Self-Distillation MAE – were applied to pre-train on the unlabeled COVID-CTset dataset. All self-supervised methods outperformed direct training, demonstrating enhanced performance through self-supervised pre-training. Models have been tested with varying positive and negative sample ratios to evaluate robustness under real-world class imbalances. Experiments showed that self-supervised learning provides a more robust feature representation foundation. Comparison with CNNs pre-trained on ImageNet revealed that self-supervised learning provides higher accuracy. While supervised pre-training

performed slightly better with balanced data, self-supervised learning exhibited stronger generalization to highly imbalanced scenarios.

In Conclusions section, the dissertation summarizes the results that hold both academic value and practical applicability.

The dissertation is formatted in accordance with the requirements of the Order of the Ministry of Education and Science of Ukraine No. 40 of January 12, 2017 "On Approval of the Requirements for Dissertation Formatting".

Publication of the results of the dissertation.

The scientific results of the dissertation are presented in 4 scientific publications by the candidate, including: 4 articles in periodic scientific journals indexed in the Web of Science Core Collection and Scopus databases, of which 1 article is in journals ranked in the first quartile (Q1) according to the SCImago Journal and Country Rank or Journal Citation Reports classification;

Additionally, the results of the dissertation were presented at 3 scientific professional conferences.

Thus, the scientific results described in the dissertation are fully reflected in the candidate's scientific publications.

Shortcomings and comments on the dissertation.

1. While the developed methods have led to enhancements in performance metrics, the value of improvement observed is less than 0.3% comparing the advancements of state-of-the-art techniques. There is a question of balancing algorithm complexity and accuracy improvement achieved.
2. From the results presented, it is not clear, whether the improvement is achieved by parameter optimization (masking ratio and mixup factor) or by the novel methods.
3. A significant limitation of the current study is the lack of experimental validation on large-scale datasets, such as ImageNet, which are critical for assessing the scalability and generalizability of the proposed approaches.
4. The math expression of the Loss function (2.8), (2.12), (2.14), (2.15) should be without 'min'. I assume that the author intended to specify the optimization criterion.
5. There are following imperfections in formatting the dissertation:
   - The reference order is not clear. It is not in order of appearance of links in the text, alphabet order, or chronological order. The requirements (item 11) provide no other ordering of sources for dissertation formatting.
   - According to the requirements for dissertation formatting (item 9), references must be to all publications listed in the abstract, and all of these publications should be in the 'References' section of the dissertation.

- The page 49 has the confusing wording "graphically depicted in an equation (2.1)".

From my personal standpoint, the expressed remarks are not decisive and do not diminish the overall scientific novelty and practical significance of the results and do not affect the positive evaluation of the dissertation.

Conclusion about the dissertation.

It could be argued that the dissertation of Xu Jiashu, submitted for earning the degree of Doctor of Philosophy, on the topic "Research and Development of Self-Supervised Visual Feature Learning Based on Neural Networks" is conducted at a high scientific level, does not violate the principles of academic integrity, and is a completed scientific research. The collective theoretical and practical results solve a scientific task of significant importance for the field of image recognition software. The dissertation work, in terms of relevance, practical value, and scientific novelty, fully meets the requirements of the current legislation of Ukraine as provided in paragraphs 6 – 9 of the "Procedure for awarding the degree of Doctor of Philosophy and cancellation of the decision of a one-time specialized academic council of a higher education institution, scientific institution on awarding the degree of Doctor of Philosophy," approved by the Resolution of the Cabinet of Ministers of Ukraine dated January 12, 2022, No. 44.

Xu Jiashu deserves to be awarded the degree of Doctor of Philosophy in the field of knowledge 12 Information Technologies with the specialty 121 Software Engineering.

Reviewer:

D. Sc., Professor of the Department
of Computer Science and Software Engineering
Igor Sikorsky Kyiv Polytechnic Institute — Inna Stetsenko

Seal

«21» 02 2024 year

Підпис Стецюк
ЗАСВІДЧУЮ
Відділ кадрів і архівної справи