FEEDBACK

by the official opponent on the dissertation work of
Xu Jiashu
on the topic «Research and Development of Self-Supervised Visual Feature Learning Based on Neural Networks»,
submitted for the degree of Doctor of Philosophy
in the field of knowledge 12 Information Technologies
with the specialization 121 Software Engineering

Relevance of the dissertation topic

The dissertation topic, "Research and Development of Self-Supervised Visual Feature Learning Based on Neural Networks," is highly relevant in computer vision. This thesis addresses the challenge of efficiently training deep learning algorithms without relying on large datasets with extensive annotations. Exploring self-learning algorithms introduces two novel approaches: the Mixup Feature and the Denoising Distillation Masked Autoencoder. These methods significantly augment the broader family of self-learning algorithms. They reduce dependence on labeled data while achieving accuracy comparable to supervised methods for learning visual features. These algorithms have tremendous potential in practical application, especially in analyzing medical images where acquiring a large volume of labeled medical images is often challenging. Given the global shortage of medical experts and the increasing demand for diagnostic services, there is an urgent need for accurate interpretation of medical images with minimal human intervention. This work could contribute to developing cost-effective pre-training schemes that eliminate the need for expensive and labor-intensive data labeling.

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:

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

A masked autoencoder model for self-supervised learning is proposed, featuring novel noise suppression and self-distillation mechanisms. 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.

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 increases the robustness of self-supervised learning models when working with unbalanced data sets.

Thus, in the dissertation work, 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 dissertation work of the candidate Xu Jiashu corresponds to the Higher Education Standard for the specialty 121 – Software Engineering in terms of competencies GC01 (General Competency 01), GC02, GC03, GC04, and SC03 (Specialized Competency 03).

The dissertation is a completed scientific work demonstrating the candidate's 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 work of Xu Jiashu is the result of the candidate's independent research and does not contain elements of falsification, compilation, fabrication, plagiarism, and borrowing. Other authors' ideas, results, and texts properly reference the relevant source.

Language and Style of Presentation of Results

The dissertation is written in English.

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

In the introduction, the PhD candidate highlights the increasingly significant role of Self-Supervised Learning (SSL) in deep learning, especially for tasks where labeled data is scarce, such as in the analysis of medical images. This research is dedicated to advancing SSL for image understanding – a field within computer vision that presents significant challenges and remains underexplored. The introduction initially emphasizes the importance of SSL and its application both in natural language processing and computer vision. The primary goal is to develop a self-directed learning system designed to extract visual features without labeled data, with a particular focus on applications in medical imaging. Furthermore, it outlines critical tasks, including developing innovative SSL algorithms, creating efficient model architectures, defining self-learning tasks and loss functions, applying transfer

learning, and evaluating model performance. The introduction concludes with a summary of its contributions and the scientific novelty it represents. Special emphasis is placed on the potential application of these findings in medical image classification, demonstrating how this research contributes to new insights and methodologies in computer vision and self-supervised learning.

The first chapter provides a comprehensive overview of the evolution of self-learning algorithms in visual feature learning. While the methodologies selected can no longer be considered competitive in the current context, these early self-regulation methods have laid the foundation upon which modern strategies are built. These algorithms can be divided into four main types: contrastive learning, masked image modeling, self-distillation, and canonical correlation analysis. This section offers a detailed review of various methodologies, starting from early methods and concluding with the latest approaches based on Vision Transformers (ViT). It meticulously examines how these methods collectively contribute to the advancement of self-supervised learning of visual representations. This thorough analysis highlights the historical progress in this field and establishes a foundation for understanding current and future trends in self-supervised learning through computer vision.

Chapter 2 presents two self-learning algorithms focused on modeling masked images, marking progress in learning from visual features. The first algorithm, Mixup Feature, introduces a novel task of reconstructing a mixture of traditional image features such as Sobel, HOG, and LBP as a target for the masked autoencoder. Using mixed features allows for the perception of more complex visual representations. The second algorithm, Denoising Self-Distillation Masked Autoencoder, combines self-distillation with masked autoencoders for effective denoising. It considers both pixel-level image restoration and feature-level regression. Gaussian noise is randomly added to image regions as a denoising task for the student network. The teacher network guides the student through exponential moving average parameter updates. An asymmetric decoder further enhances efficiency. The loss function balances reconstruction losses at the pixel level and alignment losses at the feature level.

In Chapter 3, comprehensive experiments were conducted to test the two proposed self-learning algorithms – Mixup Feature and Denoising Distillation Masked Autoencoder. For the Mixup Feature model, it was pre-trained on the CIFAR-10, CIFAR-100, and STL-10 datasets. Subsequent evaluation included linear probing and full fine-tuning. The results indicated that mixing traditional features such as Sobel, HOG, LBP enhances performance compared to individual features. An unnormalized combination of features provided the best performance when set as the target. Ideally, a masking coefficient within the 40-60% range is desirable. A mixing coefficient λ in the range of 0.2-0.3 yields the best results for the HOG-Sobel method. This method matches or exceeds MAE and surpasses contrastive learning. For the

Denoising Self-Distillation Masked Autoencoder, the model adopted ViT as the backbone network and underwent pre-training on the same three datasets. Removing randomly added blocks of Gaussian noise served as the pretext task. Full fine-tuning and linear probing tests demonstrated performance comparable to MAE and MaskFeat over 500 epochs, which is faster than the 1600 epochs for MAE. This strategy surpasses the fundamental approach of contrastive learning MoCo v3. The study confirmed the combined effect of pixel reconstruction loss and feature regression loss, demonstrating the effectiveness of the proposed framework. In conclusion, comprehensive experiments demonstrated the efficiency of the two proposed self-learning algorithms on three different datasets compared to existing methods. The results affirmed the innovative improvement of modeling masked images with mixed targets and MAE with denoising distillation. Further work could assess the applicability of these methods to larger datasets.

Chapter 4 explores the application of self-learning models to classify CT scans. This chapter examines how self-learning models are trained on unlabeled CT images and addresses critical issues. Three self-learning methods – MAE, Mixup Feature, and Denoising Self-Distillation MAE – were applied for pre-training on the unlabeled COVID-CTset dataset. The reconstructed images on retained data showed that the model captured the global anatomical structure but struggled with fine tissue textures. However, the efficiency of reconstruction does not necessarily correlate with learning ability. The pre-trained models were then fine-tuned on two smaller labeled datasets for classification. All self-supervised methods outperformed direct training, demonstrating improved performance due to self-supervised pre-training. Mixup Feature achieved results comparable to MAE, highlighting its potential in the medical field. The models were tested with varying ratios of positive to negative samples to assess robustness to real-class imbalances. Self-supervised pre-training stabilized performance with increasing imbalance, while direct training showed a more significant decline. This suggests that self-learning provides a more reliable foundation for feature representation. Comparison with CNNs pre-trained on ImageNet showed that self-learning maintained higher accuracy even at an extreme imbalance of 8:1. While supervised pre-training showed slightly better results on balanced data, self-learning demonstrated more robust generalization for highly unbalanced scenarios. Thus, this work showcased the promise of self-directed pre-training for enhancing CT scan analysis and reliable classification of medical images.

In the conclusions, the candidate summarizes the results, which possess 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 dated January 12, 2017, No. 40 "On Approval of the Requirements for Dissertation Formatting".

Publication of the results of the dissertation work

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, among 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 dissertation results were validated at 3 scientific professional conferences.

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

Shortcomings and comments on the dissertation work

There are several comments regarding the dissertation work:

1. The work does not cover models, methods, technologies, processes, and ways of developing and maintaining software and ensuring its quality, constituting the theoretical content of the subject area of specialty 121 Software Engineering.

2. Chapter 1, which takes up almost 30% of the dissertation's content, is overloaded with descriptions of well-known algorithms.

3. The characteristics of the proposed methods were studied using three different datasets. Obviously, this is insufficient to consider the validation of the developed software as qualitative.

4. A significant limitation of the current study is the lack of experimental verification on large-scale datasets, such as ImageNet, which is critically important for assessing the scalability and generalizability of the proposed approaches.

5. The text of the dissertation contains orthographic and stylistic inaccuracies, as well as text formatting errors.

I believe that the expressed remarks are not decisive, do not diminish the overall scientific novelty and practical significance of the results, and do not affect the positive evaluation of the dissertation work.

Conclusion about the dissertation work.

I consider that the dissertation work of the candidate for the degree of Doctor of Philosophy, Xu Jiashu, 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 completed scientific research. The collective theoretical and practical results solve a scientific task of significant importance for Self-Supervised Visual Feature Learning. 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.

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

Official opponent:
Professor of the Software Engineering Department at Odesa Polytechnic National University, Doctor of Technical Sciences, Professor — Vira LIUBCHENKO

Seal

February 19, 2024

Особистий підпис Viza Liubchenko
засвідчую
Ст. інспектор ВК
ВІДДІЛ КАДРІВ