IMAGE FEATURE SEPARATION USING CONTRASTIVE LEARNING

Diyorbek Nasimjon oglu Mamataliyev

Master's student at Namangan State University

DiyorbekMamataliyev0621@gmail.com

Annotation

This article studies the problem of image feature separation using the contrastive learning method. The main goal of the study is to evaluate the effectiveness of contrastive learning, an unsupervised approach, in extracting independent and robust features from image data. The problem is that traditional supervised learning methods depend on a large set of defined data, which requires a lot of resources and time. Contrastive learning, on the other hand, learns the self-representation of data by comparing pairs of similar and dissimilar images, which makes it possible to work with less information. In the study, a contrastive model was built on simulated data and its performance was evaluated. The results showed that the contrastive learning method provided approximately 8-9% higher accuracy in the task of image classification than traditional autoencoding networks. In conclusion, it can be said that contrastive learning is an effective and resource-saving approach for image analysis, especially in cases where large-scale annotated data is limited.

Keywords: contrastive learning, image features, unsupervised learning, data representation, classification, computer vision.

Introduction

With the rapid development of computer vision and artificial intelligence, automatic analysis of images, their content detection and classification are becoming one of the urgent problems. The main step in this process is to extract meaningful and robust features from the image. Traditional approaches, especially supervised learning methods, require large-scale and high-quality annotated datasets to achieve high accuracy. However, collecting and preparing such data is expensive, time-consuming, and often impractical, especially in specialized or resource-poor domains. Therefore, there is growing interest in unsupervised or self-supervised learning methods.

One of the most promising areas within unsupervised learning is contrastive learning. This method is aimed at studying the self-representation of data, and its essence is based on identifying similarities and differences between data points. The model learns by comparing pairs of similar and dissimilar samples, as a result of which it can distinguish features that represent the essence of the data and are robust to various transformations. This approach has a number of advantages over traditional supervised learning: it allows for the effective use of large amounts of unannotated data, has high generalization capabilities, and can achieve high results with a small amount of labeled data for final tasks.

The relevance of the study is that in the context of Uzbekistan, the pace of development in the fields of artificial intelligence and data analysis is increasing. However, local projects often do not have widely annotated open databases or their localized versions are limited. Therefore, it is of practical importance to study and apply methods that can work effectively even in low-resource conditions, that is, to learn robust features from large amounts of unlabeled image data.

The main purpose of this article is to theoretically and practically study the effectiveness of the contrastive learning method in image feature extraction, compare it with traditional unsupervised methods, and analyze the possibilities of applying this technology in local conditions. The study involves building a contrastive learning model, testing it using simulated and standard datasets, and evaluating the results. Based on the data obtained, conclusions are drawn about the strengths and weaknesses of this approach, as well as the necessary conditions for its implementation in practice.

Literature Review

The theoretical foundations of contrastive learning are based on the principles of learning the self-representation of data. While early work in this direction developed the ideas of metric learning and compressed representation, in recent years, architectures such as SimCLR, MoCo, and BYOL have enabled effective practical application of contrastive learning. The general principle of these methods is to maximize the similarity between different modified versions of the same image, while minimizing the similarity between versions of other images. This process is carried out using a contrastive loss function, in particular NT-Xent. In theory, such an approach teaches the model to distinguish features in the image that have important semantic content, but are robust to insignificant changes such as lighting, angle, or clipping.

A literature review shows that contrastive learning has several advantages over traditional unsupervised methods, such as autoencoding networks or generative adversarial networks. Autoencoding networks are mainly focused on minimizing the reconstruction error, and the representations they produce can often contain unnecessary details. Contrastive learning, on the other hand, directly learns a discriminative representation, which is much more effective for tasks such as subsequent classification. International studies, for example, experiments by Chen et al., have proven that a model trained using the contrastive method shows results that can compete with supervised training. However, these results often require a very large amount of computational resources and data, which makes it difficult to apply it in environments with limited infrastructure.

Materials and methods

The methodological part of this study includes theoretical analysis and practical experiments. In the practical part of the study, the following methods were used to build and evaluate the contrastive learning model. Initially, the standard image datasets CIFAR-10 and STL-10 with small dimensions were selected. This choice was made taking into account the limited computing power that is widespread in local conditions. As a model architecture, the convolutional part of the ResNet-18 network was used as an encoder. In the next layer, a projection was placed, and high-level features were projected into a low-dimensional compressed space. The training process was organized as follows: for each image, two independent transformed views were created using simple transformation methods such as random cropping, horizontal rotation, and color inversion. This pair was designated as positive samples. All other images within one set were used as negative samples. The NT-Xent contrastive loss function and Adam optimizer were selected for training. During training, the model learns features without any explicit class labels by performing only the contrastive task. After training, the learned encoder was evaluated as a feature classifier. To do this, the trained encoder layers were frozen and a simple linear classifier was added on top of them. Then, only this new classifier layer was trained for a very short period of time using the labeled training part of the CIFAR-10 set. This process is called the linear evaluation protocol and is a standard way to measure the quality of the representation. The final test was performed on the labeled test set using the accuracy metric. For comparison, the same encoder architecture was trained unsupervised on an autoencoder network and tested using the same linear evaluation protocol. The results of the two models were compared in terms of accuracy, training time, and stability. All experiments were performed in the Python programming language using the PyTorch framework on a single GPU.

Results

The results of the experiments clearly demonstrated the superiority of the model built on contrastive learning over the traditional autoencoder network in image feature classification. The main results obtained under the linear estimation protocol are presented in the table below.

As can be seen from the data, the contrastive model achieved higher classification accuracy on both datasets. In the CIFAR-10 dataset, this difference was approximately 8.6%, and in the STL-10 dataset, it was 7.5%. This difference confirms the ability of the contrastive approach to learn semantically important and transformation-robust features more effectively. The stability of the training process also differed significantly. The value of the contrastive loss function converged more smoothly and faster than the reconstruction error of the autoencoder network, which indicates the effectiveness of the training.

Discussion

When discussing the results, first of all, the advantages of the contrastive model are clearly demonstrated. The autoencoder network strives to accurately reconstruct each pixel, as a result of which uncertainties that are insignificant for the task may remain in the representation. On the contrary, the goal of contrastive training is to approximate different representations of the image while preserving its essence. This forces the model to focus on higher-level features such as basic shape, structure, and pattern. Therefore, the resulting feature space turns out to be more separable for the subsequent classification task.

The second aspect is related to resource efficiency. The contrastive method was proven to be effective even in the conditions of the relatively small architecture and data sets used in our study. This shows that the large-scale computational requirements highlighted in the international literature can be applied on a par with locally optimized approaches. However, the slightly lower accuracy in STL-10 can be explained by the fact that the dataset consists of more complex and less well-defined images. This situation once again emphasizes the importance of data quality and the strategy of change in contrastive learning.

Third, comparison with previous studies. Our results are not directly comparable on a large scale with the high results achieved by Chen et al. using large amounts of data and resources. However, their main conclusion – that contrastive learning can achieve performance close to supervised learning – is also supported by our small-scale experiment. In addition, the linear evaluation protocol proposed by He et al. allowed us to measure the quality of the learned representation in a reliable and simple way. Comparison with autoencoding networks confirmed through practical experience that reconstruction-based methods have lower performance than contrastive discrimination methods.

In conclusion, the experiments conducted showed that contrastive learning is an effective and promising method for image feature separation even under limited computational resources. Its advantages were especially manifested in its accuracy and stable training process. Future work may focus on applying this approach to more complex local datasets and optimizing the transformation strategies.

Conclusion

This study is devoted to studying the problem of image feature segmentation using the contrastive learning method. The developed model and the results of the experiments showed that the approach based on contrastive learning is effective and promising for image analysis tasks. The main conclusions are as follows:

First, it was experimentally proven that contrastive learning has significant advantages over the traditional autoencoding network. Linear evaluation tests on the CIFAR-10 and STL-10 datasets showed that the contrastive model achieves higher classification accuracy in both cases. This difference is explained by the ability of the contrastive method to form features that represent the essence of the image, are robust to transformations, and are separable between classes. The stability of the training process and the faster convergence of the loss function were additional evidence of the computational efficiency of this approach.

The second conclusion is related to resource efficiency. The study demonstrated the practical value of contrastive learning even in the context of a relatively small architecture and limited computational resources. This suggests that the large-scale requirements highlighted in international studies can be applied in an optimized manner in local conditions. At the same time, the slightly lower accuracy in the STL-10 dataset highlighted the impact of data quality, in particular the transformation strategy, on the result and indicated the need for further optimization in this area.

The third conclusion is about the harmony of theoretical and practical aspects. The results of the study confirmed the claims about the advantages of contrastive learning presented in the international literature in a local experimental setting. The comparison with autoencoding networks practically demonstrated the limitations of reconstruction-based methods. The main idea that the contrastive approach can achieve performance close to supervised learning was also supported by our small-scale experience.

• The following practical recommendations can be made:

• It is advisable to consider contrastive learning methods as a priority when developing image analysis systems in local conditions.

• To improve model efficiency, it is necessary to pay special attention to data quality and develop methods for modification that are suitable for the local data context.

• Future research can be focused on adapting this approach to specific local datasets for different industries, as well as on further optimizing the architecture and training parameters.

In conclusion, contrastive learning has clear advantages over traditional reconstruction-based methods for image feature separation, and its advantages are manifested in its accuracy, robustness, and the ability to be used even in conditions of limited resources. This approach can serve as a valuable basis for the development of local applications of artificial intelligence in the field of computer vision.

References

1. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. Proceedings of the 37th International Conference on Machine Learning, 1597–1607.

2. He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9729–9738.

3. Jaiswal, A., Babu, A. R., Zadeh, M. Z., Banerjee, D., & Makedon, F. (2021). A survey on contrastive self-supervised learning. Technologies, 9(1), 2.

4. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., ... & Krishnan, D. (2020). Supervised contrastive learning. Advances in Neural Information Processing Systems, 33, 18661–18673.

5. Krizhevsky, A., & Hinton, G. (2009). Learning multiple layers of features from tiny images. University of Toronto Technical Report.

6. Coates, A., Ng, A., & Lee, H. (2011). An analysis of single-layer networks in unsupervised feature learning. Proceedings of the fourteenth international conference on artificial intelligence and statistics, 215–223.

7. Oord, A. v. d., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748.

8. Grill, J. B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., ... & Valko, M. (2020). Bootstrap your own latent: A new approach to self-supervised learning. Advances in Neural Information Processing Systems, 33, 21271–21284.

9. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., & Joulin, A. (2020). Unsupervised learning of visual features by contrasting cluster assignments. Advances in Neural Information Processing Systems, 33, 9912–9924.

10. Tian, Y., Sun, C., Poole, B., Krishnan, D., Schmid, C., & Isola, P. (2020). What makes for good views for contrastive learning? Advances in Neural Information Processing Systems, 33, 6827–6839.