EasyChair Preprint

№ 15263

Is Self-Supervised Learning a Surrogate of Supervised Learning?

Usman Khalid and Mehmet Kaya

October 18, 2024

Is self-supervised learning a surrogate of supervised learning?

Usman Khalid
Dept. of Computer Engineering
Firat University
Elazig /Turkey
loftyusman@gmail.com

Mehmet Kaya
Dept. of Computer Engineering
Fırat University
Elazig / Turkey
kaya@firat.edu.tr

***Abstract*— Over the decades in the sphere of deep learning and machine learning, supervised learning has stood to be the anchor but the enormous amount of unannotated data, high cost in the annotation process, lengthy cycle involved in annotating the data, and the need for experts have been a drawback. This drawback brought semi-supervised learning into the limelight yet still semi-supervised needs a portion of annotated data. This expensiveness involved in getting the correct annotated data has brought a nascent self-supervised learning. Self-supervised learning learns useful information called pretext from the vast amount of data that is used on the downstream task. This emerging strategy has been a topic for research. The use of unannotated data to achieve supervised learning has brought the question if self-supervised is a surrogate for supervised learning. In this work, we reviewed the work of researchers to tackle the answer of whether these two strategies should be a surrogate or synergized by comparing their accuracy and their robustness to attack or detect out-of-distribution. Just like supervised learning faces problems with annotations, self-supervised learning also calls for a good pretext to be used on its downstream task. In this work, we recommended to conclude on the question asked all factors must be taken into account.**

Keywords*— *self-supervised, supervised, semi-supervised, robust, out-of-distribution pretext, downstream.

I. INTRODUCTION

Over the decades in the sphere of deep learning and machine learning supervised learning has stood out to be the anchor. Supervised learning is the ability of a model to learn insight from an annotated dataset. Fig.1[4] is an example of an annotated data. This popularity of supervised learning has been a topic area for researchers [2] and has been the most important [1] strategy in machine learning. Supervised learning cannot function without labelled data which is one drawback of the concept. This drawback comes as the result of unannotated data is substantially easier [3] to come by. To bridge the problem between labelled and unlabelled data gave birth to semi-supervised learning. Semi-supervised learning is a strategy where both annotated and unannotated data are used in the training process of a model, making it placed in between supervised and unsupervised learning [5]. Fig.2[8] shows an example of semi-supervised data. In 1998 [3] became one of the dominant early works that contributed to the development of this concept (semi-supervised learning) also the work of [6] proposed that "ideally we only need one labelled example per component". This strategy has helped to curb the headache of relying solely on labelled data yet still, there is the need for sampled labelled data for semi-supervised learning to fully function.

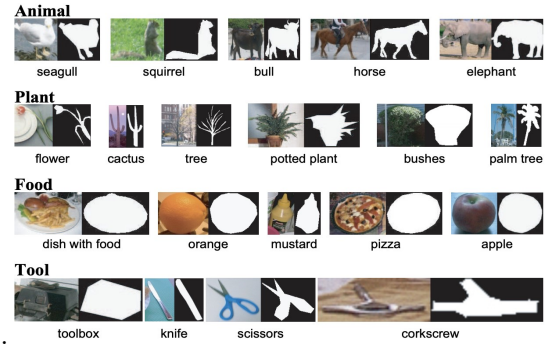

Figure.1. An annotated dataset [4]

The cost involved in labelling unannotated data, wrong annotating of data [7] and also the need for specialists to correctly label the data for it to be used is been a stumbling block for both semi-supervised and supervised learning. The challenges faced by compiling annotated data have then brought the question can researchers, machine learning and deep learning engineers rely solely on unannotated data [7] in the supervised learning concept? There comes the concept of self-supervised learning. Self-supervised aims at learning features from the unlabelled data (pretext [16]) which is then used on the downstream task. Self-supervised learning has then gained acceptance over the years and has become one of the fastest growers among researchers especially in the health sectors and areas where acquisition and annotation of data is difficult. Thus, this concept has then mitigated the need for over-reliance on annotated data.

"Most of human and animal learning is unsupervised learning. If intelligence was a cake, unsupervised learning would be the cake, supervised learning would be the icing on the cake, and reinforcement learning would be the cherry on the cake." a quote made by Yann LeCun. Today the one concept associated with unsupervised learning (self-supervised learning) has become the entire topic of discussion tending to relegate the icing on the cake. It is therefore not surprising when the French computer scientist Yann André LeCun went on to describe this concept as the future of machine learning [28].

In this work, we are enthralled to discuss and review the similarities and dissimilarities as to how the two concepts (supervised and self-supervised) can be:

1. Synergized to achieve a higher performance in the classification paradigm [23], and object detection [24,25].
2. And whether self-supervised learning is a surrogate for supervised learning in the field of computer vision.

3. And the limitations that come with the two concepts.
4. And whether in the areas where there is an abundance of annotated data the concept of self-supervised learning should be relegated.

by reviewing the works done by researchers on the topic.

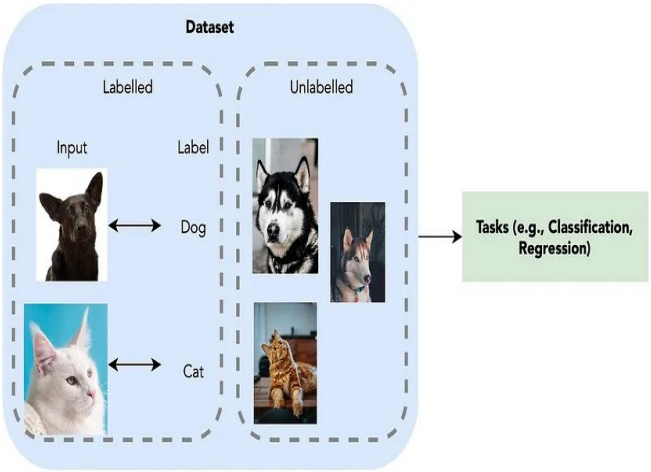

Figure. 2. A semi-supervised dataset [8]

II. RELATED WORKS

A. Self-Supervised learning

Self-supervised learning strategy makes the model learn patterns and information from the data called pretext [16] or auxiliary [13] task by turning the unsupervised data into supervised which is then trained on the main task called downstream task. Fig.3 [9] shows the pretext and downstream task. Self-supervised learning has contributed much to the field of Natural Language processing by producing many state-of-the-art pretext models which include Bert [11], Word2Vec [10], Glove [12] etc. This concept has then been leveraged to perform in computer vision, audios-visuals [34] by making use of a large amount of unannotated data at hand, and also areas where acquisition of data is scarce [14], and the need for specialists [15] in the annotation process is both expensive [14,15] and time-consuming, especially in the health sectors.

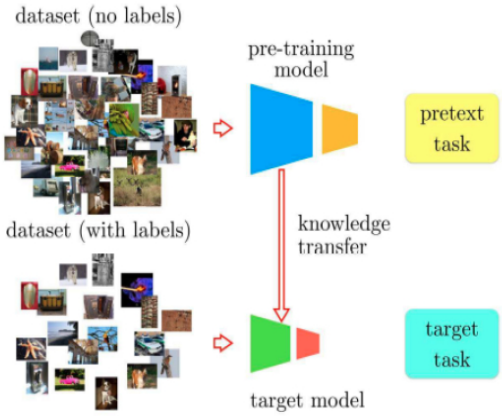

Figure. 3 A pretext and a downstream task [9]

B. Proposed Strategies By Researcher

Strategies used across self-supervised learning and early works are categorized into contrastive learning [17], generative learning [17], innate relationship [17, 19], self-predictive [17,21] and also the work of [18] proposed Local and Global Representation (LoGo) as a new strategy. Among these strategies, the two most commonly used are generative and contrastive learning [29].

- Generative learning: The ideal of the generative model gained popularity through the advent of GAN [33], and auto encoders [32]. The combination of the principle of Generative and self-supervised learning is Generative self-supervised learning. In this strategy, the pretext job is to generate some part of the data or generate the entire data, by doing so the model(pretext) learns insight or useful information which can then be fine-tuned to perform the downstream task. The work of [17] made mention of many works that used generative self-supervised learning to achieve many downstream tasks such as predicting post-traumatic stress disorder, breast cancer classification etc.

- Innate relationship: Innate learning: Innate relationship is whereby the model's main focus is on learning patterns and information through hand-crafted tasks [17]. The hand-crafted task is achieved through direct manipulation of the data such as rotation, zoom rotation, etc. The work [19,20] described learning features through rotation. The learned features can then be fine-tuned to perform the downstream task although this comes with its limitations as indicated in [17].
- Contrastive SLL: The term contrastive learning means learning the similarities (positive sample) and dissimilarities (negative sample) between the data. The combination of the idea of self-supervised and contrastive learning gives contrastive self-supervised learning. Contrastive SSL gained popularity when the framework proposed by [30] improved previous state-of-arts. The Contextual meaning of the data(image) does change after alteration as proposed in the work of [17]. In contrastive SSL the aim is to minimize the similarities distance and maximize the dissimilarities distance. The work of [31] suggested that contrastive representatives can help in clone and bug detection. This concept has been used in many frameworks like Simple Contrastive Learning of Representations (SimCLR [30]), Momentum Contrast (MoCo [35]) and SimSiam [36].
- Self-predictive: In self-prediction, the input data is augmented by hiding some part of the data and making the model predict the hidden part (pretext), by learning to predict the hidden part the model learns insight features from the data. This idea can be compared to the concept of masking in Natural Processing Language (NLP) [22], Graph Neural Networks (GNN) [21] and by removing the color

of the image and making the model predict the original color of the image, this concept is called Colorization [26,27]. The work of [17], suggested that the pretext obtained from self-predictive strategies can be fine-tuned on many downstream tasks. The work of [37] also emphasized the need for correlations between augmentations.

III. COMAPRING SUPERVICED AND SELF-SUPERVICED LEARNING

In this work, our focus is on the most important concept to evaluate the performance of a model which is the accuracy and the model's robustness to attacks and uncertainties and also its sensitivity to imbalanced data. We compared works done by researchers concerning the above-mentioned concepts. As discussed, self-supervised learning tends to curb the overreliance on annotated data, the cost and the fear of wrongful annotation of data which is related to supervised learning. The learning process involves the two differences during the final or output layers but looks similar in the mid or hidden layers [51].

A. Accuracy: Although self-supervised learning has achieved much improvement in many areas and researchers are constantly improving on it. Comparing just accuracy with supervised learning the work of [39] has proven that fully supervised learning outperforms self-supervised learning and most of the work has no huge impact on accuracy. The work of [48] indicated the underperformance of self-supervised learning to supervised learning when working on ImageNet. However, this might not always be the case since other works say the contrary [49].

B. Robustness: When a model can perform under all circumstances, handling all kinds of noise, data uncertainties, and data imbalance and be able to generalise and fit in all domains then we could say that the model is robust. To be able to test how well a model can perform in all aspects and to evaluate the model's robustness, the concept of out-of-distribution (OOD) detection and adversarial attacks are used.

- Adversarial Attack: An adversarial attack is to Intentionally perturb input data to force the model to make mistakes in its decision-making. This small perturbation makes it difficult to deploy models to production, especially in the areas where mistakes cannot be afforded. An example of NLP is by adding additional text or synonyms which slightly distort the meaning. Fig-4 shows an example of an adversarial attack taken from [40]. The works of [38,39] suggested self-supervised learning can stand against robustness to adversarial attacks, particularly [39] compared the result to traditional supervised learning which self-supervised had the upper hand also the work of [38] indicated that using self-supervised learning improved black box attacks.

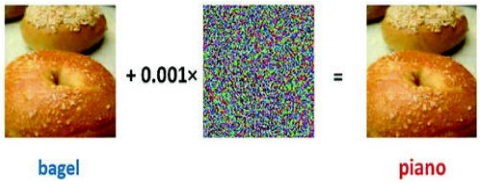

Figure. 4. An example of an adversarial attack

- Out-of-Distribution (OOD) Detection: A good model in production should be able to detect and handle data out of its training data thus, the model should be able to detect unfamiliar and actual data. It is important to detect these unfamiliar data; this is useful to prevent the model from making wrongful decisions. The work of [41] gave an excellent example related to out-of-distribution. Self-supervised learning can learn certain useful information during the pretext face which might not be useful in its downstream. These learned features become used during OOD detection. The work of [42] compared a model's robustness when encountering with imbalanced data and OOD using supervised and self-supervised learning. The work proved how self-supervised learning performs well under imbalanced data. [39] indicated by rotation self-supervised learning can detect OOD. The work of [43,44,45] all used self-supervised learning in detecting out-of-distribution. Fig-5 [47] indicates a model which was trained to detect a dog and a cat but is faced with an image of a dolphin.

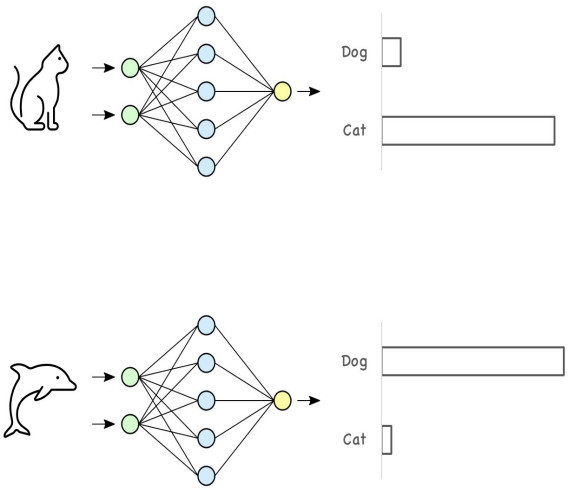

Figure. 5 An out-of-distribution problem [47]

IV. PRETEXT AND DOWNSTREAM TASK

These two concepts have been used throughout this work, and it is there necessary to classify the two terms in the field of supervised and self-supervised learning. As discussed in the early works, Pretext's job is to learn useful information from the unlabeled data which is then fine-tuned or pre-trained on a specific or main task called downstream task. A robust model from self-supervised learning requires a good pretext model. The work of [37] introduced a pretext which leveraged invariant representations called PIRL. [53] has

discussed several performing methods regarding this concept. Although self-supervised learning models have proven to be useful in many aspects as discussed and pretext works with no annotation of the input data the downstream task still requires a target data for the specific task to be done.

V. DISCUSSIONS AND CONCLUSION

Although self-supervised learning has proved to be useful in many aspects including computer vision and natural processing language with a good track record of having the upper hand during adversarial attacks, out-of-distribution, solving the problems involved in the annotation of data and some cases outperforming the fully supervised learning. Yet self-supervised learning has its drawback which involves. The need for sophisticated pretext algorithm [50]. Just like supervised learning cannot function without correctly annotated data same as self-supervised learning cannot without a pretext which has insight or useful information about the task to be done. There is a need to combine the strategies to leverage their positive effects just like the work of [52] combined self-supervised and semi-supervised. The work [39] suggested that self-supervised learning should not be viewed as autonomy but rather combine self-supervised learning and fully supervised to get the best out of the two concepts.

In this work, we discussed their differences, similarities and areas where they outperformed by examining the works done by researchers. To certainly conclude the question, it would be necessary to perform a substantial experiment using all the strategies involved in self-supervised learning, the same metrics and hyper-parameters for both fully supervised learning, semi-supervised learning and self-supervised learning to conclude the question. We hope to extend our research by performing a comprehensive practical implementation of all the recommendations mentioned

REFERENCES

[1] Cunningham, P., Cord, M., & Delany, S. J. (2008). Supervised learning. In Machine learning techniques for multimedia: case studies on organization and retrieval (pp. 21-49). Berlin, Heidelberg: Springer Berlin Heidelberg.

[2] Nasteski, V. (2017). An overview of the supervised machine learning methods. Horizons. b, 4, 51-62.

[3] Blum, A., & Mitchell, T. (1998, July). Combining labelled and unlabeled data with co-training. In Proceedings of the eleventh annual conference on Computational learning theory (pp. 92-100).

[4] Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: a database and web-based tool for image annotation. International journal of computer vision, 77, 157-173.

[5] Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. Machine learning, 109(2), 373-440.

[6] Cholaquidis, A., Fraiman, R., & Sued, M. (2020). On semi-supervised learning. TEST, 29(4), 914-937.

[7] Xiaojin, Z. (2006). Semi-supervised learning literature survey. Semi-supervised learning Literature Survey, Technical report, Computer Sciences, University of Wisconsin-Madisoa.

[8] Cheng, T. (2021). Supervised, Semi-Supervised, Unsupervised, and Self-Supervised Learning. Towards Data Science. https://towardsdatascience. com/supervised-semi-supervised-unsupervised-and-self-supervised-learning-7fa79aa9247c.

[9] Noroozi, M., Vinjimoor, A., Favaro, P., & Pirsiavash, H. (2018). Boosting self-supervised learning via knowledge transfer. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 9359-9367).

[10] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 26.

[11] Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).

[12] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[13] Ohri, K., & Kumar, M. (2021). Review on self-supervised image recognition using deep neural networks. Knowledge-Based Systems, 224, 107090

[14] Hou, Y., & Sang, Q. (2023, March). Boosting Ultrasonic Image Classification via Self-Supervised Representation Learning. In 2023 3rd International Conference on Computer, Control and Robotics (ICCCR) (pp. 116-120). IEEE.

[15] Huang, S. C., Pareek, A., Jensen, M., Lungren, M. P., Yeung, S., & Chaudhari, A. S. (2023). Self-supervised learning for medical image classification: a systematic review and implementation guidelines. NPJ Digital Medicine, 6(1), 74.

[16] Krishnan, R., Rajpurkar, P., & Topol, E. J. (2022). Self-supervised learning in medicine and healthcare. Nature Biomedical Engineering, 6(12), 1346-1352

[17] Huang, S. C., Pareek, A., Jensen, M., Lungren, M. P., Yeung, S., & Chaudhari, A. S. (2023). Self-supervised learning for medical image classification: a systematic review and implementation guidelines. NPJ Digital Medicine, 6(1), 74.

[18] Zhang, T., Qiu, C., Ke, W., Süsstrunk, S., & Salzmann, M. (2022). Leverage your local and global representations: A new self-supervised learning strategy. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16580-16589).

[19] Feng, Z., Xu, C., & Tao, D. (2019). Self-supervised representation learning by rotation feature decoupling. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10364-10374).

[20] Li, X., Hu, X., Qi, X., Yu, L., Zhao, W., Heng, P. A., & Xing, L. (2021). Rotation-oriented collaborative self-supervised learning for retinal disease diagnosis. IEEE Transactions on Medical Imaging, 40(9), 2284-2294.

[21] Schlichtkrull, M. S., De Cao, N., & Titov, I. (2020). Interpreting graph neural networks for nlp with differentiable edge masking. arXiv preprint arXiv:2010.00577.

[22] Kaneko, M., & Bollegala, D. (2022, June). Unmasking the mask–evaluating social biases in masked language models. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 11, pp. 11954-11962).

[23] Breiki, F. A., Ridzuan, M., & Grandhe, R. (2021). Self-supervised learning for fine-grained image classification. arXiv preprint arXiv:2107.13973.

[24] Zhao, X., Pang, Y., Zhang, L., Lu, H., & Ruan, X. (2022, June). Self-supervised pretraining for RGB-D salient object detection. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 3, pp. 3463-3471).

[25] Lee, W., Na, J., & Kim, G. (2019). Multi-task self-supervised object detection via recycling of bounding box annotations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4984-4993).

[26] Zhang, R., Isola, P., & Efros, A. A. (2016). Colorful image colorization. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands,

October 11-14, 2016, Proceedings, Part III 14 (pp. 649-666). Springer International Publishing.

[27] Larsson, G., Maire, M., & Shakhnarovich, G. (2016). Learning representations for automatic colorization. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14 (pp. 577-593). Springer International Publishing.

[28] Hinton G, LeCunn Y, Bengio Y. AAAI'2020 keynotes turing award winners event. https://www.youtube.com/watch?v=UX8OubxsY8w

[29] Lilit Yolyan Review on Self-Supervised Contrastive Learning Brief introduction and overview of self-supervised contrastive learning https://towardsdatascience.com/review-on-self-supervised- contrastive-learning-93171f695140

[30] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PMLR.

[31] Yang, S., Gu, X., & Shen, B. (2022, May). Self-supervised learning of smart contract representations. In Proceedings of the 30th IEEE/ACM International Conference on Program Comprehension (pp. 82-93)

[32] Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

[33] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

[34] Terbouche, H., Schoneveld, L., Benson, O., & Othmani, A. (2022). Comparing Learning Methodologies for Self-Supervised Audio-Visual Representation Learning. IEEE Access, 10, 41622-41638.

[35] He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9729-9738).

[36] Chen, X., & He, K. (2021). Exploring simple siamese representation learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 15750-15758).

[37] Misra, I., & Maaten, L. V. D. (2020). Self-supervised learning of pretext-invariant representations. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 6707-6717).

[38] Kim, M., Tack, J., & Hwang, S. J. (2020). Adversarial self-supervised contrastive learning. Advances in Neural Information Processing Systems, 33, 2983-2994.

[39] Hendrycks, D., Mazeika, M., Kadavath, S., & Song, D. (2019). Using self-supervised learning can improve model robustness and uncertainty. Advances in neural information processing systems, 32.

[40] Chen, P. Y., Zhang, H., Sharma, Y., Yi, J., & Hsieh, C. J. (2017, November). Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models. In Proceedings of the 10th ACM workshop on artificial intelligence and security (pp. 15-26).

[41] Yang, J., Zhou, K., Li, Y., & Liu, Z. (2021). Generalized out-of-distribution detection: A survey. arXiv preprint arXiv:2110.11334

[42]] Liu, H., HaoChen, J. Z., Gaidon, A., & Ma, T. (2021). Self-supervised learning is more robust to dataset imbalance. arXiv preprint arXiv:2110.05025.

[43] Vyas, A., Jammalamadaka, N., Zhu, X., Das, D., Kaul, B., & Willke, T. L. (2018). Out-of-distribution detection using an ensemble of self supervised leave-out classifiers. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 550-564).

[44] Mohseni, S., Pitale, M., Yadawa, J. B. S., & Wang, Z. (2020, April). Self-supervised learning for generalizable out-of-distribution detection. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 04, pp. 5216-5223).

[45] Rafiee, N., Gholamipoor, R., Adaloglou, N., Jaxy, S., Ramakers, J., & Kollmann, M. (2022, September). Self-supervised anomaly detection by self-distillation and negative sampling. In International Conference on Artificial Neural Networks (pp. 459-470). Cham: Springer Nature Switzerland

[46] Sehwag, V., Chiang, M., & Mittal, P. (2021). Ssd: A unified framework for self-supervised outlier detection. arXiv preprint arXiv:2103.12051.

[47] Mun Hou's Detecting Out-of-Distribution Samples with kNN https://blog.munhou.com/2022/12/01/Detecting%20Out-of-Distribution%20Samples%20with%20Knn/

[48] Tomasev, N., Bica, I., McWilliams, B., Buesing, L., Pascanu, R., Blundell, C., & Mitrovic, J. (2022). Pushing the limits of self-supervised ResNets: Can we outperform supervised learning without labels on ImageNet?. arXiv preprint arXiv:2201.05119.

[49] Haja, A., Brouwer, E., & Schomaker, L. (2023). Self-Supervised Versus Supervised Training for Segmentation of Organoid Images. arXiv preprint arXiv:2311.11198.

[50] Zhao, Z., Alzubaidi, L., Zhang, J., Duan, Y., & Gu, Y. (2023). A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations. Expert Systems with Applications, 122807.

[51] Grigg, T. G., Busbridge, D., Ramapuram, J., & Webb, R. (2021). Do Self-Supervised and Supervised Methods Learn Similar Visual Representations?. arXiv preprint arXiv:2110.00528.

[52] Zhai, X., Oliver, A., Kolesnikov, A., & Beyer, L. (2019). S4l: Self-supervised semi-supervised learning. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1476-1485).

[53] Albelwi, S. (2022). Survey on self-supervised learning: auxiliary pretext tasks and contrastive learning methods in imaging. Entropy, 24(4), 551.