BABEŞ-BOLYAI UNIVERSITY
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Unsupervised representation learning and feature fusion in supervised tasks. Applications in natural sciences

PhD Thesis summary

PhD student: Albu Alexandra-Ioana
Scientific supervisor: Prof. Dr. Czibula Gabriela

2023

Keywords: Representation Learning, Autoencoders, Contrastive Learning, Feature Fusion, Learning with Label Noise.

Contents

Thesis table of contents

List of publications

The ranking of publications was performed according to the CNATDCU (National Council for the Recognition of University Degrees, Diplomas and Certificates) standards applicable for doctoral students enrolled after October 1, 2018. All rankings are listed according to the classification of journals[1] and conferences[2] in Computer Science.

Publications in Web of Science - Science Citation Index Expanded

[9] **Alexandra-Ioana Albu**, Maria-Iuliana Bocicor, Gabriela Czibula. *MM-StackEns: A new deep multimodal stacked generalization approach for protein-protein interaction prediction*, Computers in Biology and Medicine, Vol. 153, 106526, 2023 (**AIS Quartile Q2** according to JCR 2022 - **1st position**, WoS category: Engineering, Biomedical).

Rank A, 8 points.

[12] **Alexandra-Ioana Albu**, Gabriela Czibula, Andrei Mihai, Istvan-Gergely Czibula, Sorin Burcea, and Abdelkader Mezghani. *NeXtNow: A convolutional deep learning model for the prediction of weather radar data for nowcasting purposes.* Remote Sensing 14, no. 16 (2022): 3890 (**AIS Quartile Q2** according to JCR 2021)

Rank B, 1 point.

[25] Gabriela Czibula, Andrei Mihai, **Alexandra-Ioana Albu**, Istvan-Gergely Czibula, Sorin Burcea, Abdelkader Mezghani. *AutoNowP: An approach using deep autoencoders for precipitation nowcasting based on radar echo prediction.* Mathematics 9(14), Special Issue on "Computational Optimizations for Machine Learning", 1653, 2021 (**IF Quartile Q1** according to JCR 2020).

Rank A, 2 points.

[23] Gabriela Czibula, **Alexandra-Ioana Albu**, Maria Iuliana Bocicor, and Camelia Chira. *AutoPPI: An ensemble of deep autoencoders for protein–protein interaction prediction.* Entropy 23, no. 6 (2021): 643 (**IF Quartile Q2** according to JCR 2020)

Rank B, 2 points.

[1] https://uefiscdi.ro/premierea-rezultatelor-cercetarii-articole
[2] http://portal.core.edu.au/conf-ranks/

Publications in Web of Science - Conference Proceedings Citation Index

[8] **Alexandra-Ioana Albu**. *Temporal ensembling deep k-nearest neighbours for learning with noisy labels*. 31st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2023), **accepted for publication**.

Rank B - CORE 2021, 4 points.

[7] **Alexandra-Ioana Albu**. *Improving radar echo extrapolation models using autoencoder-based perceptual losses*. 27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2023), **accepted for publication**.

Rank B - CORE 2021, 4 points.

[6] **Alexandra-Ioana Albu**. *An approach for predicting protein-protein interactions using supervised autoencoders*. 26th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES2022), Procedia Computer Science, Volume 207, 2022, pages 2023–2032.

Rank B - CORE 2021, 4 points.

[5] **Alexandra-Ioana Albu**. *Towards learning transferable embeddings for protein conformations using variational autoencoders*. 25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES2021), Procedia Computer Science, Volume 192, 2021, pages 10–19.

Rank B - CORE 2021, 4 points.

[10] **Alexandra-Ioana Albu**, Gabriela Czibula. *Analysing protein dynamics using machine learning based generative models*. IEEE 14th International Symposium on Applied Computational Intelligence and Informatics, SACI 2020, Timisoara, Romania, pages 000135-000140, IEEE, 2020.

Rank D - CORE2020, 1 point.

[11] **Alexandra-Ioana Albu**, Gabriela Czibula. *Learning with label noise through pairwise distance agreement between supervised and self-supervised representations*. Submitted.

Papers published in international journals and proceedings of international conferences

[4] **Alexandra-Ioana Albu**, Alina Enescu, Luigi Malagò. *Tumor detection in brain MRIs by computing dissimilarities in the latent space of a variational autoencoder*. Proceedings of the Northern Lights Deep Learning Workshop, Norway, pages 1–6, 2020.

Rank D, 1 point.

Other publications

[2] **Alexandra-Ioana Albu**, Alina Enescu, Luigi Malagò. *Detection of tumours in brain MRIs with variational autoencoders*. ECML PKDD 2020 Workshop Machine Learning for Pharma and Healthcare Applications, 2020.

Rank D, 0 point.

[3] **Alexandra-Ioana Albu**, Alina Enescu, Luigi Malagò. *Improved slice-wise tumour detection in brain MRIs by computing dissimilarities between latent representations.* 2020 KDD Workshop on Applied Data Science for Healthcare, 2020.

Rank D, 0 point.

Publications score: 31 points.

Introduction

The aim of this thesis is to explore the use of unsupervised representation learning and feature fusion techniques in order to improve the performance of supervised deep learning approaches.

Deep Learning (DL) defines a general framework for training powerful neural network models capable of learning high-level representations for the input data, by automatically extracting the information relevant for a specific task [37]. Despite the fact that supervised DL methods have achieved remarkable performance in numerous tasks, their success depends on the availability of large, accurately annotated data sets. However, in many real world settings, labelled data is scarce, noisy, or imbalanced, and therefore, the generalization of DL methods can be significantly affected [16, 34, 82]. The unsupervised learning paradigm, on the other hand, is able to learn information about the structure of the data without using annotations. Therefore, it has the potential to provide a useful tool for complementing labels and may subsequently enhance the performance of supervised models. Unsupervised representation learning has emerged as one of the most important subfields of unsupervised DL. It focuses on learning high-level representations that characterize the input samples and which can be useful for downstream tasks [15, 34]. Widely studied representation learning models include various types of autoencoders, contrastive learning methods and self-supervised language models [34, 52]. The study of the advantage brought by representation learning in classification or regression tasks has been a topic of interest in the literature, the main research directions focusing on pre-training [34, 77], anomaly detection models [18] or semi-supervised techniques [28, 48]. Through the models introduced in this thesis, we explored new directions and identified new tasks in which unsupervised representation learning may benefit supervised learning. Apart from the way in which representations are learned (through supervised or unsupervised methods), the literature also studied methods for optimally fusing the representations obtained from different components or branches of the models [27, 68]. Due to the close connection between the methods for learning and for fusing representations and to their complementary nature, we explored the two topics in conjunction with each other.

The first problem approached in our thesis consists of classification with noisy labels. Learning robust models using data sets having incorrect labels is challenging. Nevertheless, the problem has a major practical importance, originating from the fact that in real-world settings the labelling process is prone to noise, either because of human mistakes or because of incorrect automatic labelling [56, 79]. Since deep neural networks can learn to fit random annotations [86], the development of algorithms that can generalize to new data instances even if trained with unreliable or incorrect samples is essential. In this context, we developed two methods which aimed to exploit unsupervised information from the data in order to guide the supervised learning process.

Our first approach, *TE-kNN* [8], has the goal of identifying samples in a data set that are likely to have correct labels. These selected samples are further used in the training of neural networks. The proposed *TE-kNN* is similar to existing sample selection strategies which compute an agreement score of a sample with its nearest neighbours in feature space for each epoch. This score is further used to decide if that sample is correctly annotated [13, 60].

Unlike other approaches, *TE-kNN* adopts a temporal ensembling [50] strategy that takes into account the scores previously obtained by the sample during training. In this context, we investigated the benefit of an initialization scheme for the scores using the neighbours given by an autoencoder trained on the data. We showed, through experiments on three image data sets, that our method is superior to several related approaches, including the original nearest neighbours strategy.

While effective in low or moderate noise regimes, sample selection approaches are not that suitable for severe label noise settings, since the set of clean, selected labels is very small [67]. In order to address this noise scenario, we proposed a new regularization method, *DIAG* [11], which leverages self-supervised contrastive learning with the aim of obtaining representations that are less affected by label noise. The proposed procedure includes a new regularizer that matches pairwise distances computed in the supervised feature space to the distances obtained using a self-supervised model. Our approach was motivated by the intuition that learning representations that agree with both supervised and self-supervised training regimes should diminish the effect of incorrect annotations. Experiments performed on both synthetic and realistic label noise pointed out that *DIAG* outperforms multiple literature approaches on this task.

The two proposed methods are complementary and address different types of noise regimes. While *TE-kNN* is more well suited for low and moderate-noise settings, the advantage brought by *DIAG* is more pronounced in high noise contexts. With these two contributions, we aimed to highlight the potential of representation learning in training neural networks using different types and amounts of label noise.

The second direction followed in our research consisted of the study of two applied research topics relevant for the field of natural sciences. We studied an important area from biology, namely protein-protein interaction (PPI) prediction and one relevant topic from the meteorological domain, represented by weather nowcasting, for which we hypothesized that unsupervised representation learning and feature fusion could be explored in order to improve the performance of supervised DL techniques.

Proteomics consists of the study of proteins, which are large molecules that influence all the processes taking place in living organisms. The identification of PPIs is an important problem in proteomics, having the potential to advance research in protein function prediction and drug design. Since experimental methods for detecting PPIs are very complex and imply high costs [69], there is a need for developing reliable computational approaches for this task. The problem can be framed as a binary classification task in which the instances are pairs of proteins. The interactome can also be viewed as a graph whose nodes are represented by proteins and whose edges are given by interactions. A challenge for PPI prediction methods is represented by the difficulty of accurately detecting interactions for testing proteins that are not included in the training graph, as compared to the easier task of identifying new interactions for previously encountered proteins [31, 47, 61]. Due to this generalization problem, we focused on leveraging representation learning methods in order to design PPI prediction models. Moreover, the PPI task can be seen as a classification problem for pairs of samples, thus our second focus was on the development of better fusion techniques for protein pairs.

In [23] we proposed an anomaly detection-inspired binary classification model, *AutoPPI*, for identifying PPIs. The model trains two autoencoders: one on interacting proteins and the other on non-interacting proteins, and detects the class of a pair of proteins using the autoencoder which more accurately reconstructs the pair. In order to better reconstruct the protein pairs, we designed two new siamese autoencoder architectures, which make use of feature fusion modules. *AutoPPI* outperformed the majority of related work approaches on the considered data sets (13 cases out of 16 comparisons).

Although *AutoPPI* provided good performance on the evaluated data sets, it has diffi-

culties in predicting interactions for proteins which are not included in the training graph. Our second approach [6] aimed to address this limitation by proposing a method based on *supervised autoencoders* [44, 51]. These are neural networks that jointly optimize a classification branch and a reconstruction branch, the latter adding a regularization effect to the network [44, 51]. Our proposed approach is a two-stage classification method encompassing a denoising pretraining stage and a joint, supervised autoencoder training stage. We showed the good results obtained by our approach in comparison to multiple related work methods.

With our third approach, *MM-StackEns* [9], we aimed to further improve the performance on unseen proteins by learning more comprehensive representations for characterizing PPI networks. Unlike our first two approaches that use only protein sequences information, *MM-StackEns* is an ensemble procedure consisting of two structurally different neural networks, aimed at exploiting multimodality in PPI networks, namely a sequence component and a graph component. Additionally, we leveraged self-supervised representation learning in the context of PPI identification through embeddings learned using a deep pre-trained language model. We showed that such embeddings can provide improved generalization for interactions of out-of-graph proteins. Another contribution of this work is the design of a new feature fusion module for protein pairs that can be easily integrated in both the sequence and graph components of *MM-StackEns*. The proposed approach outperformed related work methods in the great majority of evaluation scenarios. To summarize, this approach exploited both feature fusion techniques as well as self-supervised representation learning in order to improve the identification of PPIs.

The second application studied in this thesis is represented by weather nowcasting using radar data. With the recent increase in severe weather phenomena, the ability to accurately predict weather evolution is becoming of greater and greater importance. Radar reflectivity is one of the products most commonly used by meteorologists for detecting weather phenomena. Weather nowcasting can be viewed as a spatio-temporal prediction task in which we aim to learn a mapping from a sequence of past radar values measurements to the measurements corresponding to one or multiple time frames in the future [63]. Radar data is high-dimensional and noisy, with errors caused by radar measurements that can detect various non-meteorological objects from the environment. In addition, meteorological data is highly imbalanced, since severe events are rare [36]. Nevertheless, accurate prediction of such extreme events is crucial for a weather forecasting system. Given the rich spatial structure of the data, convolutional architectures are a very promising avenue of research in this field. However, models trained with traditional loss functions tend to output blurry predictions [42]. These characteristics of the weather nowcasting problem motivated us to study feature fusion techniques in designing convolutional architectures and to investigate the potential of autoencoders in designing classification models and new loss functions.

In [25] we have introduced *AutoNowP*, which is an approach for the weather nowcasting problem, modelled as a binary classification problem at the pixel-level. We learn to classify whether a data point will have a value above or below a threshold using the neighbours of the point at a previous timestamp. This model is similar to the *AutoPPI* method in terms of training procedure and classification methodology. *AutoNowP* differs from *AutoPPI* in the fact that it uses class-specialized loss functions for each autoencoder. This modification has the goal of introducing an additional inductive bias for the two models, in a scenario in which the inputs characterizing the two classes are very similar/overlapping. In addition, in this study we evaluated the proposed model in an imbalanced setting. We compared *AutoNowP* with several machine learning baselines using radar data from Romania and Norway and outlined the competitive results obtained by our method.

Our second approach for weather nowcasting, *NeXtNow* [12], is a deeper convolutional architecture, which aims to be more scalable than *AutoNowP*, by modelling the radar data at the map-level instead of the pixel-level. The proposed model is a convolutional architecture

built using *ResNeXt* blocks [83]. The study aimed to investigate whether feature fusion operations as the one contained in *ResNeXt* blocks can be useful for weather nowcasting. Our model was evaluated using radar data sets from both Romania and Norway.

While the previous two studies focused on the development of new models and architectures, in our third study we shifted our attention towards the design of improved loss functions with the goal of addressing the blurriness problem of DL weather nowcasting models. To this end, we designed in [7] a new family of perceptual losses [46] using autoencoders as feature extractors instead of the commonly used deep classifiers (eg: VGG) pre-trained on ImageNet. Experiments on radar reflectivity data collected by MET Norway showed that our method is able to provide sharper predictions than two classical loss functions, while achieving a small improvement in terms of root-mean-square error. The proposed approach has the additional advantage of being more lightweight than perceptual losses employing VGG networks.

The chosen applied topics both process imbalanced and noisy labels, which traditional supervised methods struggle to effectively exploit. This motivated us to investigate approaches based on representation learning and feature fusion. Moreover, by designing and applying such methods in these two fields, we aimed to show the suitability of combining concepts from supervised learning with unsupervised representation learning and feature fusion on sequential data (proteins) and data having both spatial and temporal structure (weather data).

The content of this thesis is structured as follows. Chapter 1 presents an overview of DL, including the main building blocks used in designing our proposed models and an in-depth literature review of the studied problems.

Chapter 2 presents our proposed approaches for improving classification models in the presence of label noise using representation learning. Section 2.1 presents our first approach, *TE-kNN* [8], which is a sample selection method. In Section 2.2, our *DIAG* [11] regularization method aimed at improving the robustness of deep neural networks to label noise is introduced.

Chapter 3 presents our original contributions in designing improved PPI prediction models using techniques from representation learning and feature fusion. Section 3.1 presents our *AutoPPI* [23] approach, alongside the proposed siamese autoencoder architectures. In Section 3.2 we detail our proposed approaches using supervised autoencoders for PPI prediction [6] and evaluate their generalization to proteins outside the training graph. Section 3.3 presents our third contribution in the PPI prediction task, the *MM-StackEns* [9] multimodal stacked generalization approach for PPI, alongside an extensive experimental evaluation of the proposed model on multiple data sets and interaction types.

Chapter 4 contains our original contributions developed for designing nowcasting models. Section 4.1 presents the *AutoNowP* binary classification model [25] built using autoencoders and its potential in nowcasting tasks. The *NeXtNow* convolutional architecture [12] is presented in Section 4.2 as an illustration of the usefulness of feature fusion techniques in the nowcasting domain. Section 4.3 presents our approach for improving the visual quality of predicted radar maps through designing autoencoder-based perceptual losses [7].

Finally, the last section outlines the conclusions of this thesis and directions of extending the proposed models.

***

The author acknowledges financial support from the NO Grants 2014-2021, under Project contract no. 26/2020.

Chapter 1

Background

This chapter presents an overview of the main theoretical aspects covering the DL models used as components of the classification models developed in this thesis (Section 1.1). Subsequently, the ML and DL approaches proposed in the literature for the applied problems studied in our research, namely PPI prediction (Section 1.2) and weather nowcasting (Section 1.3), are reviewed.

1.1 Deep Learning: theoretical notions and building blocks

Deep learning is represented by the study of deep neural networks [37]. The manner in which these models are trained is influenced by the *learning paradigm* that is being used. In *supervised learning*, a model is trained with a collection of data-annotation pairs in order to associate data instances to a discrete label (in classification problems) or a continuous value (in regression problems) [37]. However, for most problems, a large data set having high quality annotations is needed in order to properly learn this association between data instances and their labels, which may not be feasible in many real world applications [34].

A different paradigm, which can discover patterns and structure in the input data without using annotations is represented by *unsupervised learning* [37]. A subdomain of unsupervised learning gaining significant attention recently is represented by the *self-supervised* learning paradigm. The main characteristic of *self-supervised* learning is the fact that it generates a pretext task starting from the available unlabelled data and trains a model on this pretext task [34]. In the following, some of the main types of unsupervised representation learning methods are briefly presented.

Autoencoders are neural networks which learn high-level representations for input samples by being trained to reconstruct the instances using the learned representations. A *denoising autoencoder* learns to recreate the input from a corrupted version of itself [37].

The contrastive learning framework comprises a series of self-supervised techniques that obtain representations for data samples such that similar samples are clustered together, while dissimilar ones are well separated. This is achieved by defining pairs of similar samples using augmentations [21].

Another successful use of the self-supervised paradigm is represented by language models, used to obtain word embeddings [34]. Recently, contextualized models, such as ELMo [62] and BERT [29], have been introduced, which take into account the context (entire sentence) of a word when computing its embedding. The pretext task which is typically defined is the prediction of a token in a sequence given the previous tokens or the prediction of a masked part of the sequence from the remaining sequence.

The representations learned in an unsupervised way through these models can be useful in supervised contexts. The main directions pursued so far in the literature focused on unsupervised pre-training, followed by supervised training. This can be performed either

using the fixed learned features or by fine-tuning these representations on the target task [33, 34, 69]. These approaches can be beneficial if we do not have many labelled instances for the supervised task. Another important application is represented by anomaly detection [18], for which autoencoders are frequently used. Joint supervised and unsupervised training methods, which helped in preventing overfitting, have also been proposed [51].

A situation in which traditional supervised learning does not provide good generalization and which is frequently encountered in practice is represented by the presence of incorrect annotations, or *label noise*. Various approaches have been developed with the goal of reducing the impact of incorrect annotations [67]. Motivated by the observation that deep learning models tend to overfit noisy labels, regularization techniques have been studied [14, 43, 45, 56, 70, 82, 88]. Another strategy for learning with label noise is represented by the identification of samples which are most likely to be correct and then training using only these selected samples [13, 60].

1.2 Protein-Protein interaction prediction

The first applied task investigated in this thesis is represented by PPI prediction. Accurate detection of PPIs can facilitate a better understanding of protein functions and biological mechanisms [69].

Numerous machine learning approaches for determining PPIs from protein sequences have been proposed. Earlier research focused on classical machine learning techniques [22, 38]. In addition, neural networks have been studied [20, 89]. A series of approaches have investigated the usefulness of autoencoders as feature extractors in the context of PPI identification [69, 80, 81]. Other pursued directions were represented by ensemble methods [19, 54], graph neural networks [57, 85] and multimodal approaches [55, 87].

Although the performance of PPI methods has constantly improved over the years, a few works have pointed out the difference in performance encountered when testing PPI approaches on pairs of proteins which are part of the training graph compared to the performance obtained for pairs formed of previously unseen proteins [31, 57, 61]. With the aim of providing a comprehensive evaluation protocol for PPI systems, Park and Marcotte [61] proposed three methods of creating test sets: one in which both proteins in the pair have been seen in the training phase, but in interactions with other proteins (denoted in their paper as the **C1** class), one in which only one protein has not been seen (**C2**-type interactions) and one in which neither of the proteins appears in the training set interactions (**C3** class). Their study emphasized the need to include all three types of testing data sets when evaluating PPI prediction methods, since in real-world scenarios only approximately 19.2% of the interactions represent C1-type interactions, while C2 links amount for 49.2% and C3-type interactions represent 31.6% [61].

1.3 Weather nowcasting

This section introduces the second application from the field of natural sciences that we explored in this thesis, namely *weather nowcasting*.

Weather nowcasting is the prediction of the weather evolution for a short period of time in the future, consisting of lead times of at most two to six hours. Being able to accurately predict weather phenomena can be of crucial importance for human activity, especially in the case of severe and rapidly changing weather events [1].

One of the main data sources for weather nowcasting is represented by radar reflectivity data, due to the known existent methods for estimating rainfall from reflectivity data. Reflectivity is a product measured by radars periodically, which consists of numerical values gathered for a disk corresponding to the geographical location surrounding the radar [36].

Since the radar data has both spatial and temporal structure, which both play an important role in nowcasting, the problem can be modelled as a spatio-temporal prediction task [64].

Various deep learning weather nowcasting methods have been proposed, which can be broadly split in two categories, defined by the manner in which the temporal dimension is being modelled. The first category of approaches makes use of fully convolutional networks exclusively. The proposed models are based on U-Net [1, 75], Xception [65], 3D convolutions [40] or causal convolutions [17]. The second category of approaches consists of models based on convolutional LSTMs (ConvLSTMs) [64, 71, 84] which represent customizations of LSTMs suitable for modelling spatio-temporal data.

A general drawback of deep learning nowcasting models is the fact that they are not able to accurately predict severe events, represented by high reflectivity values, because of the high degree of data imbalance [36]. Another limitation observed by multiple studies is that the predictions of weather nowcasting models tend to be much smoother and lose detail in comparison to the original radar maps [42, 74, 76]. This issue is caused by pixel-wise losses such as the Mean squared error (MSE), which assume that the data is sampled from a Normal distribution [58]. The main directions pursued for improving the visual quality of the predictions were the use of Generative Adversarial Networks [73] and perceptual losses [76, 78], as well as combinations of the two [42]. The idea behind perceptual losses is that of computing a distance between the predicted and true image not directly in pixel space, but in a feature space that distills higher level characteristics of the images, such as style, content or textures [30].

Chapter 2

New approaches for classification with noisy labels using representation learning

The current chapter presents our contributions focused on developing new methods for robust learning in the presence of noisy labels by leveraging unsupervised information in the data. We propose two methods, *TE-kNN* (presented in Section 2.1) and *DIAG* (presented in Section 2.2), falling in the sample selection and regularization methods categories, respectively, which are two of the most important research directions in learning with label noise. The methods presented in this chapter are original contributions introduced in two studies [8, 11].

2.1 Temporal ensembling deep k-nearest neighbours for learning with noisy labels

One of the prominent research directions for learning with noisy labels is represented by sample selection methods, which aim to identify instances with correct annotations in order to use only the selected, reliable samples during training [13, 39]. A technique used in several works for sample selection identifies clean samples as the instances which agree in terms of label with their k nearest neighbours [13, 35, 60]. The neighbours are usually retrieved by computing distances between the representations obtained from the penultimate layer in the neural network [35, 60]. We argue, however, that considering only the neighbours obtained for the current epoch can be sub-optimal, due to the incorrect annotations and the dynamics of the training process which can lead to inaccurate neighbour sets.

For addressing this limitation, we proposed in [8] a deep k-NN sample selection method, which adopts a temporal ensembling [50] procedure for computing the scores used for deciding if a sample has a correct label. More specifically, our approach computes a weighted average between previous scores of a sample and the score obtained during the current epoch, thus benefiting from the model's experience. We investigate two initializations for the scores: one in which all initial scores are zero and one in which the initial scores are computed using the neighbours given by an autoencoder trained on the data.

The experimental evaluation was performed on the SVHN [59], CIFAR-10 and CIFAR-100 [49] data sets using synthetically generated label noise. The results outlined the superiority of our approach over several baselines, including the original k-NN strategy.

2.2 Learning with label noise through pairwise distance agreement between supervised and self-supervised representations

While sample selection methods are effective in low label noise settings, such approaches would ignore the majority of the data in high noise regimes [67]. Thus, in [11], we proposed a regularization method aimed primarily at improving the robustness of neural networks trained with severe label noise. The intuition behind our proposed approach is that exploiting unsupervised information about the structure of the input data through self-supervised learning can be beneficial for supervised neural networks trained with noisy labels.

Our proposed method encourages a pairwise distance agreement between the supervised embeddings of the samples in the batch and the corresponding embeddings obtained using a self-supervised model. We hypothesized that learning representations which agree with both supervised and self-supervised training regimes should diminish the effect of incorrect annotations. The self-supervised model was represented by a contrastive learning model trained on the data.

An analysis of the related work approaches on learning with label noise showed that our approach is conceptually different from other literature approaches for learning with label noise. Experiments were performed on data sets containing synthetic noise (CIFAR-10 and CIFAR-100 [49]) and realistic noise (Animal-10N [66] and Food-101N [53]). The obtained results highlighted the good performance of our method, which achieved an improvement of up to 61.6% over the best competing method on severe label noise and of up to 1.8% for moderate noise.

2.3 Conclusions

This chapter presented our proposed approaches using representation learning for classification in the presence of noisy labels: a new sample selection method and a new regularizer. Their effectiveness in comparison with related methods was shown on multiple image data sets. We note however, that the developed approaches are general and could be applied to different types of data sets as well.

As previously outlined, the two proposed methods are conceptually complementary. Therefore, future work will study strategies for combining the two proposed approaches.

Chapter 3

Representation learning and feature fusion for predicting protein-protein interactions

This chapter presents our new approaches for the PPI task which make use of representation learning models, more specifically autoencoders and self-supervised language models. The feature fusion modules that we proposed with the aim of better capturing patterns from pairs of proteins are also described.

The models and experiments presented in this chapter have been published in [6, 9, 23].

The remaining of this chapter is organized as follows. Section 3.1 presents our *AutoPPI* approach. Section 3.2 presents our proposed approach using supervised autoencoders for PPI prediction. Section 3.3 introduces and details our third contribution in the PPI prediction task, the *MM-StackEns* multimodal stacked generalization approach for PPI prediction.

3.1 *AutoPPI*: protein-protein interaction prediction using autoencoders

The first approach that we proposed for the task of PPI prediction is a sequence-based binary supervised classifier, *AutoPPI*, published in [23]. This classifier is an anomaly detection-inspired [24, 72] approach composed of two autoencoders used for learning the characteristics of each of the two available classes: the class of proteins that interact and the class of proteins that do not interact.

Due to the fact that we train autoencoders on *pairs* of proteins, we designed two new architectures making use of intermediate fusion modules which are more suitable for paired samples than classical architectures. The first proposed architecture has a shared structure only for the encoder, while aiming to reconstruct the concatenated features of the two proteins. The second proposed architecture has a siamese structure in both the encoder and decoder.

An analysis of the literature highlighted that our approach is new in the PPI prediction literature. We evaluated the proposed method using four public data sets belonging to multiple species and the obtained results showed that *AutoPPI* is superior to the majority of the related work methods on the considered data sets (wining in 81.25% of the comparisons with the related work).

3.2 Supervised autoencoders for predicting protein-protein interactions

A limitation encountered for many PPI prediction models, including our *AutoPPI* is the sub-optimal performance on C2 and C3 interactions (i.e. interactions containing one or two proteins which are not part of the training graph).

In [6], we introduced a model aimed at addressing this limitation. The proposed method is a two-stage approach based on *supervised autoencoders*. The first stage consists of pre-training a denoising autoencoder on protein sequences. In the second stage, a classification network is attached to the pre-trained encoder. The model simultaneously learns to predict the class (interaction or non-interaction) associated to protein pairs and to reconstruct the two proteins in the pair. The choice of a supervised autoencoder was motivated by the regularization effect of adding a reconstruction branch observed in previous work [44,51]. Similarly, the denoising pre-training stage has the goal of learning robust initial representations. To the best of our knowledge, supervised autoencoders have not been used for PPI prediction so far.

We evaluated both the proposed approach and a classical *supervised autoencoder* on two public data sets using different types of interactions (C1, C2 and C3) and showed that our models obtained better results than multiple related work methods.

3.3 *MM-StackEns*: A multimodal ensemble method for predicting protein-protein interactions

Our third model, *MM-StackEns*, aims to improve upon the approaches presented in previous sections by narrowing the gap between the performance achieved for proteins inside the training graph compared to those that are not included in the training graph.

The proposed approach, published in [9], is a multimodal ensemble learning method which processes both sequence and graph information. The choice of a multimodal solution was motivated by the observation that processing single-source data, as in our previous approaches (*AutoPPI* and the supervised autoencoders), may not fully capture the diverse information present in available PPIs. We hypothesize that by aggregating the predictions obtained by processing both amino acid and graph topology information, more informative predictions can be obtained, due to the structural differences between the two models. While the graph model exploits the neighbourhood information of a node in order to infer other possible neighbours, the sequence component relies only on the sequence information of the two proteins in the pair. The predictions given by the two models are combined in a two-layer stacked generalization strategy that uses a Logistic Regression classifier. To improve individual models and to account for the uncertainty present in the protein data, we proposed to learn probabilistic representations for the proteins and we introduced a new feature fusion module for protein pairs. Another contribution of this study is that of leveraging self-supervised representation learning in the context of PPI identification, by using embeddings obtained via the ELMo deep language model [41,62], which was pretrained on large databases of protein sequences. We experimentally showed that this input data representation is particularly beneficial for predicting interactions of unseen proteins (i.e. for which no interactions have been encountered during training), as it outperforms classical protein representations computed using physico-chemical descriptors.

A comprehensive comparison with related work methods on several intra-species and inter-species data sets highlighted an improvement achieved by our *MM-StackEns* classifier over multiple related work methods and its good performance on C2 and C3 interactions.

3.4 Conclusions

This chapter presented our proposed models for the PPI prediction problem which take advantage of unsupervised representation learning and feature fusion techniques. Our proposed approaches were evaluated in a multitude of scenarios using various data sets and interaction types.

Future work will investigate ways to incorporate components that discriminate between classes in the training procedure of *AutoPPI*, since, currently, each autoencoder is trained using a single class and it has only information characteristic to that class, but no information that contrasts two distinct classes.

In the approach based on supervised autoencoders, we used architectures that reconstruct independently the two proteins in a pair. Nevertheless, other architectures and reconstruction schemes could be employed, such as the architectures proposed for *AutoPPI*, which aim to model pairs of proteins and not individual sequences. Alternative possibilities are the use of stochastic models or the design of learning schemes in which one protein from the pair is used for reconstructing the other protein. These directions of research will be studied in the future.

Regarding the *MM-StackEns* model, a possible improvement for the proteins representations can be brought by investigating newer deep language models pre-trained on protein sequences, such as ProtBert, ProtAlbert or ProtT5 [32] instead of the ELMo representation.

Chapter 4

Representation learning and feature fusion for weather nowcasting

This chapter presents our original contributions developed with the aim of leveraging representation learning and feature fusion techniques in order to build new, improved deep learning models for weather nowcasting.

The models and experiments presented in this chapter have been published in three papers [7, 12, 25].

The current chapter is structured as follows. Section 4.1 presents the *AutoNowP* binary classification model. The *NeXtNow* nowcasting model is presented in Section 4.2. Section 4.3 details our approach for designing autoencoder-based perceptual losses.

4.1 *AutoNowP*: an approach using convolutional autoencoders for radar reflectivity prediction

In [25], we proposed *AutoNowP*, an approach for nowcasting based on autoencoders. We modelled the problem as a binary classification task in which we aimed to output whether the reflectivity values at a location will be larger than a pre-specified threshold at a moment in the future, given the reflectivity values measured by the radar at a previous time moment. A binary output is relevant for weather nowcasting, since such thresholds correspond to meteorological events of different intensities [26].

Few literature approaches focused on studying representation learning in the context of weather nowcasting. In this work, we aimed to explore such a direction, by introducing an approach for nowcasting using two autoencoders, each trained on the instances from a different class. In order to encourage the two autoencoders to learn to reconstruct better the reflectivity values having the same label as the class corresponding to that autoencoder, we introduced a different weighted loss function for each of the two models. We opted for this method in order to evaluate the usefulness of the characteristics learned by an autoencoder for one class of reflectivity values in distinguishing between the two classes. Our proposed model is similar to the model we introduced in [23] for PPI prediction, with the difference that for *AutoNowP*, convolutional architectures were used in order to exploit the spatial structure of the data and weighted loss functions were designed. Thus, with this contribution, we show the potential of the autoencoder based approach previously used for PPI prediction [23] in the weather nowcasting applied domain.

We evaluated our approach using radar data from Romania and Norway and outlined its competitive performance.

4.2 *NeXtNow*: a new convolutional architecture for weather nowcasting

In [12], we investigated the potential of feature fusion methods for improving weather nowcasting models, by introducing a new convolutional model for radar reflectivity prediction, which makes use of ResNeXt blocks [83]. Our proposed model has an encoder–decoder architecture that consists of ResNeXt blocks and simple convolutions, which maps one or multiple past radar measurements to one or multiple future radar maps. Therefore, *NeXtNow* is a customized version of ResNeXt for the short-term prediction of radar data. In addition, in the study published in [12], an empirical analysis was performed with the aim of evaluating the impact of using a larger temporal context, containing several time steps in the past, on the performance of the model.

The *NeXtNow* model was evaluated using two data sets obtained from Romania and Norway. The obtained results highlighted that *NeXtNow* outperformed, in the majority of cases, a convolutional architecture [65] that was proposed for short-term radar data prediction.

4.3 Autoencoder-based perceptual losses for improving weather nowcasting models

Our third proposed approach for weather nowcasting, introduced in [7], aimed at enhancing the visual quality of the radar maps predicted by DL models. Our proposed methodology uses perceptual losses, which compute a distance between predicted and true images in a feature space obtained using deep neural networks. Perceptual losses have previously been investigated in the computer vision and nowcasting literature and are usually formed using neural network classifiers (e.g. VGGs) pretrained on ImageNet [42, 76, 78]. However, while features extracted by a VGG can accurately model natural RGB images, the extracted features may not be ideal for radar data, which has different characteristics and appearance. An additional disadvantage is given by the fact that VGGs are large-scale architectures.

In [7], we proposed an alternative for using deep pre-trained classifiers for extracting the high level representations used in the computation of the perceptual losses. Instead, we proposed to use convolutional autoencoders trained on the radar data. Our model is a more lightweight architecture compared to the classical VGG and has the advantage of not requiring any additional annotations or data. We showed that our model provides sharper predictions than classical pixel-wise losses, while achieving a small improvement in terms of root-mean-square error.

4.4 Conclusions

In this chapter, our approaches using representation learning and feature fusion for the weather nowcasting problem have been presented. A classification model, a convolutional architecture and a new type of loss function were proposed.

Future work will extend the evaluation of the three approaches in more challenging weather nowcasting scenarios, involving predictions for longer time periods in the future and covering larger geographical regions.

In addition, we plan to evaluate *AutoNowP* on nowcasting tasks which use different data models. A possibility is to use *AutoNowP* in order to predict if a severe weather alert will be issued at a future time step. Such a scenario would imply representing data instances as an entire map covered by a radar, characterized by the radar reflectivity values for that region.

NeXtNow can be extended by including modules from our architecture inside recurrent models. Additionally, more complex feature fusion techniques could be studied, such as

attentional modules [27]. Therefore, the $NeXtNow$ model can constitute a starting point for developing improved DL architectures for weather nowcasting.

Other directions of research will investigate the possibility of combining the proposed perceptual losses with adversarial training techniques in order to further improve the sharpness of the obtained predictions. In addition, future work will study the use of multiple feature maps obtained from different layers of the autoencoders in the computation of the perceptual loss. We note that even though we evaluated the proposed approach on radar reflectivity prediction, the use of autoencoders as feature extractors is general and could be applied to other image generation or prediction tasks.

Conclusions

This thesis had the main goal of investigating methods for improving supervised neural network classifiers using techniques and ideas from unsupervised representation learning and feature fusion. In particular, the problem of classification when having access to data sets containing incorrect annotations was studied. A second goal of our research was to evaluate the suitability and applicability of our proposed approaches in two research domains from the natural sciences, which have the common characteristic of involving noisy and imbalanced data and for which training a model using classical supervised techniques does not imply a good generalization on the test set.

Chapter 2 presented our approaches for improving the performance of deep neural networks trained on data sets containing label noise. We presented a sample selection approach, *TE-kNN*, which introduced a temporal ensembling procedure in order to compute more robust sample scores and a regularizer, *DIAG*, which makes use of contrastive learning representations in order to guide the training of supervised models. The good performance of both approaches and their superiority over multiple related work methods on data sets using synthetic and realistic label noise was highlighted.

In Chapter 3 we presented our approaches for tackling the PPI prediction problem. The three presented approaches included a classification model using autoencoders, *AutoPPI*, alongside which two new siamese architectures were proposed. Our second approach uses supervised autoencoders for PPI prediction and experiments on data sets with three levels of difficulty for classifying interactions showed that the proposed models achieved good results in all evaluation scenarios. The third model, *MM-StackEns*, which was built on the insights given by the first two approaches, further improved the generalization of PPI prediction models with the final aim of narrowing the gap between the performance achieved for proteins inside and outside the training graph. With this approach, we proposed to learn more expressive representations for the protein pairs, through combining sequence and graph modalities. In addition, we evaluated the performance improvement brought by the use of a self-supervised language model for representing protein sequences and we introduced a new feature fusion method for protein pairs. The proposed approach outperforms the great majority of related work approaches on multiple data sets. The *AutoPPI* and *MM-StackEns* approaches explored both representation learning methods and feature fusion modules, highlighting the complementarity of the two directions.

Chapter 4 presented the models proposed in this thesis for the weather nowcasting problem. We introduced the $AutoNowP$ autoencoder-based classification model for nowcasting. Through developing this approach, we presented a unifying perspective on the two studied applied domains from the field of natural sciences, by showing that autoencoders can be used for developing effective classifiers in both the PPI prediction and weather nowcasting tasks. Our second model, $NeXtNow$, was proposed with the aim of studying feature fusion operations and their utility in nowcasting. Our third approach for weather nowcasting uses autoencoders in designing perceptual losses which can help in making the outputs of DL nowcasting models less blurry. $AutoNowP$ and $NeXtNow$ were evaluated on radar data from both Norway and Romania, while the last approach was evaluated on a data set covering a geographical region from Norway.

Through the developed approaches, the benefit that unsupervised representation learning and feature fusion can bring to supervised learning methods was investigated in a variety of contexts. From the point of view of the employed data, the two studied research problems (PPI prediction and weather nowcasting) process data with diverse structure and characteristics. From the perspective of the developed models, throughout this thesis, representation learning and feature fusion methods have been used both in designing architectural components or classification models and as a tool for creating new loss functions.

In addition to the future work directions outlined in individual chapters, another envisaged research direction will aim to assess the generality and practical relevance of the methods proposed in Chapter 2, by evaluating them on the PPI prediction and weather nowcasting tasks.

In this thesis, the development of feature fusion modules was primarily concerned with intermediate and late-fusion techniques. Nevertheless, additional directions of future work could include the investigation of early-fusion strategies for both protein and meteorological data. Another closely related extension of our work would be the study of feature fusion techniques for multimodal data. *MM-StackEns* was our first step in this direction, but our approach combined only the predictions of the models and did not fuse intermediate multimodal representations. Such a direction is of interest, since PPIs can be characterized by multiple types of information, such as protein structures, sequences or evolutionary information [31], while in the meteorological domain, additional types of data (e.g. satellite images or orographic data) could be used in conjunction with radar maps [36].

A research direction that was touched upon but not fully explored in our work is represented by learning with imbalanced data. Future work will focus on enhancing the performance of our proposed approaches in such settings.

Another important aspect, from the point of view of using the models in real-world settings, and which will be studied in the future consists of providing interpretability for the predictions, in both PPI and weather nowcasting tasks.

Bibliography

[1] S. Agrawal, L. Barrington, C. Bromberg, J. Burge, C. Gazen, and J. Hickey. Machine learning for precipitation nowcasting from radar images. In *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, pages 1–6, 2019.

[2] A. Albu, A. Enescu, and L. Malago. Detection of Tumours in Brain MRIs with Variational AutoEncoders. In *ECML PKDD 2020 Workshop Machine Learning for Pharma and Healthcare Applications*, pages 1–6, 2020.

[3] A. Albu, A. Enescu, and L. Malago. Improved Slice-wise Tumour Detection in Brain MRIs by Computing Dissimilarities between Latent Representations. In *2020 KDD Workshop on Applied Data Science for Healthcare*, 2020.

[4] A. Albu, A. Enescu, and L. Malago. Tumor detection in brain MRIs by computing dissimilarities in the latent space of a variational autoencoder. In *Proceedings of the Northern Lights Deep Learning Workshop*, pages 1–6. Septentrio Academic Publishing, 2020.

[5] A.-I. Albu. Towards learning transferable embeddings for protein conformations using variational autoencoders. *Procedia Computer Science*, 192:10–19, 2021.

[6] A.-I. Albu. An approach for predicting protein-protein interactions using supervised autoencoders. *Procedia Computer Science*, 207:2023–2032, 2022.

[7] A.-I. Albu. Improving radar echo extrapolation models using autoencoder-based perceptual losses. In *27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems 2023*. Accepted, 2023.

[8] A.-I. Albu. Temporal ensembling deep k-nearest neighbours for learning with noisy labels. In *31st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2023)*. Accepted, 2023.

[9] A.-I. Albu, M.-I. Bocicor, and G. Czibula. MM-StackEns: A new deep multimodal stacked generalization approach for protein-protein interaction prediction. *Computers in Biology and Medicine*, page 106526, 2023.

[10] A.-I. Albu and G. Czibula. Analysing protein dynamics using machine learning based generative models. In *2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pages 000135–000140. IEEE, 2020.

[11] A.-I. Albu and G. Czibula. Learning with label noise through pairwise distance agreement between supervised and self-supervised representations. Submitted, 2023.

[12] A.-I. Albu, G. Czibula, A. Mihai, I. G. Czibula, S. Burcea, and A. Mezghani. NeXtNow: A Convolutional Deep Learning Model for the Prediction of Weather Radar Data for Nowcasting Purposes. *Remote Sensing*, 14(16):3890, 2022.

[13] D. Bahri, H. Jiang, and M. Gupta. Deep k-nn for noisy labels. In *International Conference on Machine Learning*, pages 540–550. PMLR, 2020.

[14] Y. Bai, E. Yang, B. Han, Y. Yang, J. Li, Y. Mao, G. Niu, and T. Liu. Understanding and improving early stopping for learning with noisy labels. *Advances in Neural Information Processing Systems*, 34:24392–24403, 2021.

[15] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[16] K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma. Learning imbalanced datasets with label-distribution-aware margin loss. *arXiv preprint arXiv:1906.07413*, 2019.

[17] R. Castro, Y. M. Souto, E. Ogasawara, F. Porto, and E. Bezerra. Stconvs2s: Spatiotemporal convolutional sequence to sequence network for weather forecasting. *Neurocomputing*, 426:285–298, 2021.

[18] R. Chalapathy and S. Chawla. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*, 2019.

[19] K.-H. Chen, T.-F. Wang, and Y.-J. Hu. Protein-protein interaction prediction using a hybrid feature representation and a stacked generalization scheme. *BMC bioinformatics*, 20(1):1–17, 2019.

[20] M. Chen, C. J.-T. Ju, G. Zhou, X. Chen, T. Zhang, K.-W. Chang, C. Zaniolo, and W. Wang. Multifaceted protein–protein interaction prediction based on Siamese residual RCNN. *Bioinformatics*, 35(14):i305–i314, 2019.

[21] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[22] X.-W. Chen and M. Liu. Prediction of protein–protein interactions using random decision forest framework. *Bioinformatics*, 21(24):4394–4400, 2005.

[23] G. Czibula, A.-I. Albu, M. I. Bocicor, and C. Chira. AutoPPI: An Ensemble of Deep Autoencoders for Protein–Protein Interaction Prediction. *Entropy*, 23(6):643, 2021.

[24] G. Czibula, C. Codre, and M. Teletin. AnomalP: An approach for detecting anomalous protein conformations using deep autoencoders. *Expert Systems with Applications*, 166:114070, 2021.

[25] G. Czibula, A. Mihai, A.-I. Albu, I.-G. Czibula, S. Burcea, and A. Mezghani. AutoNowP: an approach using deep autoencoders for precipitation nowcasting based on weather radar reflectivity prediction. *Mathematics*, 9(14):1653, 2021.

[26] G. Czibula, A. Mihai, and I. G. Czibula. Radrar: A relational association rule mining approach for nowcasting based on predicting radar products' values. *Procedia Computer Science*, 176:300–309, 2020. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 24th International Conference KES2020.

[27] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard. Attentional feature fusion. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3560–3569, 2021.

[28] M. Deudon. Learning semantic similarity in a continuous space. In *Advances in neural information processing systems*, pages 986–997, 2018.

[29] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2019.

[30] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE transactions on pattern analysis and machine intelligence*, 44(5):2567–2581, 2020.

[31] B. Dunham and M. K. Ganapathiraju. Benchmark Evaluation of Protein-Protein Interaction Prediction Algorithms. *Molecules*, 27(1), 2022.

[32] A. Elnaggar, M. Heinzinger, C. Dallago, G. Rehawi, W. Yu, L. Jones, T. Gibbs, T. Feher, C. Angerer, M. Steinegger, et al. ProtTrans: Towards Cracking the Language of Lifes Code Through Self-Supervised Deep Learning and High Performance Computing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[33] D. Erhan, A. Courville, Y. Bengio, and P. Vincent. Why does unsupervised pre-training help deep learning? In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 201–208. JMLR Workshop and Conference Proceedings, 2010.

[34] L. Ericsson, H. Gouk, C. C. Loy, and T. M. Hospedales. Self-supervised representation learning: Introduction, advances, and challenges. *IEEE Signal Processing Magazine*, 39(3):42–62, 2022.

[35] C. Feng, G. Tzimiropoulos, and I. Patras. Ssr: An efficient and robust framework for learning with unknown label noise. In *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*. BMVA Press, 2022.

[36] G. Franch, D. Nerini, M. Pendesini, L. Coviello, G. Jurman, and C. Furlanello. Precipitation nowcasting with orographic enhanced stacked generalization: Improving deep learning predictions on extreme events. *Atmosphere*, 11(3), 2020.

[37] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[38] Y. Guo, L. Yu, Z. Wen, and M. Li. Using support vector machine combined with auto covariance to predict protein-protein interactions from protein sequences. *Nucleic acids research*, 36(9):3025–3030, 2008.

[39] B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang, and M. Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31, 2018.

[40] L. Han, J. Sun, and W. Zhang. Convolutional neural network for convective storm nowcasting using 3-d doppler weather radar data. *IEEE Transactions on Geoscience and Remote Sensing*, 58(2):1487–1495, 2020.

[41] M. Heinzinger, A. Elnaggar, Y. Wang, C. Dallago, D. Nechaev, F. Matthes, and B. Rost. Modeling aspects of the language of life through transfer-learning protein sequences. *BMC bioinformatics*, 20(1):1–17, 2019.

[42] Y. Hu, L. Chen, Z. Wang, X. Pan, and H. Li. Towards a more realistic and detailed deep-learning-based radar echo extrapolation method. *Remote Sensing*, 14(1):24, 2021.

[43] A. Iscen, J. Valmadre, A. Arnab, and C. Schmid. Learning with neighbor consistency for noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4672–4681, 2022.

[44] D. Jarrett and M. van der Schaar. Target-Embedding Autoencoders for Supervised Representation Learning. In *International Conference on Learning Representations*, 2019.

[45] I. Jindal, M. Nokleby, and X. Chen. Learning deep networks from noisy labels with dropout regularization. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 967–972. IEEE, 2016.

[46] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016.

[47] D. Kimothi, P. Biyani, J. Hogan, and M. Davis. Sequence representations and their utility for predicting protein-protein interactions. *IEEE/ACM transactions on computational biology and bioinformatics*, 2021.

[48] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling. Semi-supervised learning with deep generative models. In *Advances in neural information processing systems*, pages 3581–3589, 2014.

[49] A. Krizhevsky, G. Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[50] S. Laine and T. Aila. Temporal ensembling for semi-supervised learning. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

[51] L. Le, A. Patterson, and M. White. Supervised autoencoders: improving generalization performance with unsupervised regularizers. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 107–117, 2018.

[52] P. H. Le-Khac, G. Healy, and A. F. Smeaton. Contrastive representation learning: A framework and review. *IEEE Access*, 8:193907–193934, 2020.

[53] K.-H. Lee, X. He, L. Zhang, and L. Yang. Cleannet: Transfer learning for scalable image classifier training with label noise. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5447–5456, 2018.

[54] F. Li, F. Zhu, X. Ling, and Q. Liu. Protein interaction network reconstruction through ensemble deep learning with attention mechanism. *Frontiers in Bioengineering and Biotechnology*, 8, 2020.

[55] L. Liu, X. Zhu, Y. Ma, H. Piao, Y. Yang, X. Hao, Y. Fu, L. Wang, and J. Peng. Combining sequence and network information to enhance protein–protein interaction prediction. *BMC bioinformatics*, 21(16):1–13, 2020.

[56] S. Liu, J. Niles-Weed, N. Razavian, and C. Fernandez-Granda. Early-learning regularization prevents memorization of noisy labels. *Advances in neural information processing systems*, 33:20331–20342, 2020.

[57] G. Lv, Z. Hu, Y. Bi, and S. Zhang. Learning unknown from correlations: Graph neural network for inter-novel-protein interaction prediction. *arXiv preprint arXiv:2105.06709*, 2021.

[58] M. Mathieu, C. Couprie, and Y. LeCun. Deep multi-scale video prediction beyond mean square error. In *4th International Conference on Learning Representations, ICLR 2016*, 2016.

[59] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng. Reading digits in natural images with unsupervised feature learning. 2011.

[60] D. Ortego, E. Arazo, P. Albert, N. E. O'Connor, and K. McGuinness. Multi-objective interpolation training for robustness to label noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6606–6615, 2021.

[61] Y. Park and E. M. Marcotte. Flaws in evaluation schemes for pair-input computational predictions. *Nature methods*, 9(12):1134–1136, 2012.

[62] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. Deep Contextualized Word Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[63] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28, 2015.

[64] X. Shi, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo. Deep learning for precipitation nowcasting: A benchmark and a new model. *arXiv preprint arXiv:1706.03458*, 2017.

[65] I. A. Socaci, G. Czibula, V.-S. Ionescu, and A. Mihai. *xnow*: A deep learning technique for nowcasting based on radar products' values prediction. In *IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI 2020)*, pages 117–122. IEEE Hungary Section, 2020.

[66] H. Song, M. Kim, and J.-G. Lee. Selfie: Refurbishing unclean samples for robust deep learning. In *International Conference on Machine Learning*, pages 5907–5915. PMLR, 2019.

[67] H. Song, M. Kim, D. Park, Y. Shin, and J.-G. Lee. Learning from noisy labels with deep neural networks: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[68] S. R. Stahlschmidt, B. Ulfenborg, and J. Synnergren. Multimodal deep learning for biomedical data fusion: a review. *Briefings in Bioinformatics*, 23(2):bbab569, 2022.

[69] T. Sun, B. Zhou, L. Lai, and J. Pei. Sequence-based prediction of protein protein interaction using a deep-learning algorithm. *BMC bioinformatics*, 18(1):1–8, 2017.

[70] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[71] C. K. Sønderby, L. Espeholt, J. Heek, M. Dehghani, A. Oliver, T. Salimans, J. Hickey, S. Agrawal, and N. Kalchbrenner. Metnet: A neural weather model for precipitation forecasting. *arXiv*, pages 1–17, 2020.

[72] M. Teletin, G. Czibula, and C. Codre. AutoSimP: An Approach for Predicting Proteins' Structural Similarities Using an Ensemble of Deep Autoencoders. In *International Conference on Knowledge Science, Engineering and Management*, pages 49–54. Springer, 2019.

[73] L. Tian, X. Li, Y. Ye, P. Xie, and Y. Li. A generative adversarial gated recurrent unit model for precipitation nowcasting. *IEEE Geoscience and Remote Sensing Letters*, 17(4):601–605, 2020.

[74] Q.-K. Tran and S.-k. Song. Computer vision in precipitation nowcasting: Applying image quality assessment metrics for training deep neural networks. *Atmosphere*, 10(5):244, 2019.

[75] K. Trebing, T. Stanczyk, and S. Mehrkanoon. Smaat-unet: Precipitation nowcasting using a small attention-unet architecture. *Pattern Recognition Letters*, 145:178–186, 2021.

[76] M. Veillette, S. Samsi, and C. Mattioli. Sevir: A storm event imagery dataset for deep learning applications in radar and satellite meteorology. *Advances in Neural Information Processing Systems*, 33:22009–22019, 2020.

[77] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(12), 2010.

[78] Y. Wang and H. A. Karimi. Perceptual loss function for generating high-resolution climate data. *Applied Computing and Intelligence*, 2(2):152–172, 2022.

[79] Y. Wang, W. Liu, X. Ma, J. Bailey, H. Zha, L. Song, and S.-T. Xia. Iterative learning with open-set noisy labels. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8688–8696, 2018.

[80] Y. Wang, Z. You, L. Li, L. Cheng, X. Zhou, L. Zhang, X. Li, and T. Jiang. Predicting protein interactions using a deep learning method-stacked sparse autoencoder combined with a probabilistic classification vector machine. *Complexity*, 2018, 2018.

[81] Y.-B. Wang, Z.-H. You, X. Li, T.-H. Jiang, X. Chen, X. Zhou, and L. Wang. Predicting protein–protein interactions from protein sequences by a stacked sparse autoencoder deep neural network. *Molecular BioSystems*, 13(7):1336–1344, 2017.

[82] X. Xia, T. Liu, B. Han, C. Gong, N. Wang, Z. Ge, and Y. Chang. Robust early-learning: Hindering the memorization of noisy labels. In *International conference on learning representations*, 2020.

[83] S. Xie, R. B. Girshick, P. Dollár, Z. Tu, and K. He. Aggregated Residual Transformations for Deep Neural Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5987–5995, 2017.

[84] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pages 802–810, 2015.

[85] F. Yang, K. Fan, D. Song, and H. Lin. Graph-based prediction of protein-protein interactions with attributed signed graph embedding. *BMC bioinformatics*, 21(1):1–16, 2020.

[86] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals. Understanding deep learning requires rethinking generalization. *International conference on learning representations*, 2017.

[87] D. Zhang and M. R. Kabuka. Multimodal deep representation learning for protein-protein interaction networks. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 595–602. IEEE, 2018.

[88] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*, 2018.

[89] L. Zhao, J. Wang, Y. Hu, and L. Cheng. Conjoint feature representation of go and protein sequence for ppi prediction based on an inception rnn attention network. *Molecular Therapy-Nucleic Acids*, 22:198–208, 2020.