

A Survey on Deep Semi-supervised Learning

Based on: <https://arxiv.org/pdf/2103.00550.pdf>

Original Authors: Xiangli Yang, Zixing Song, Irwin King, Fellow, IEEE, Zenglin Xu, Senior Member, IEEE

Date Published: 23 August 2021

Introduction

What is deep semi-supervised learning (DSSL)?

- Combo of supervised and unsupervised learning which uses a small portion of labeled examples and a large number of unlabeled data
- Model must learn and make predictions on new examples.
- The basic procedure involves using the existing labeled data to label the rest of the unlabelled data, thus effectively helping to increase the training data.

Why DSSL is needed?

- Most of the real-world use cases do not have extensive labeled data and labeling data is challenging and requires a considerable amount of resources, time, and effort.



General idea of SSL

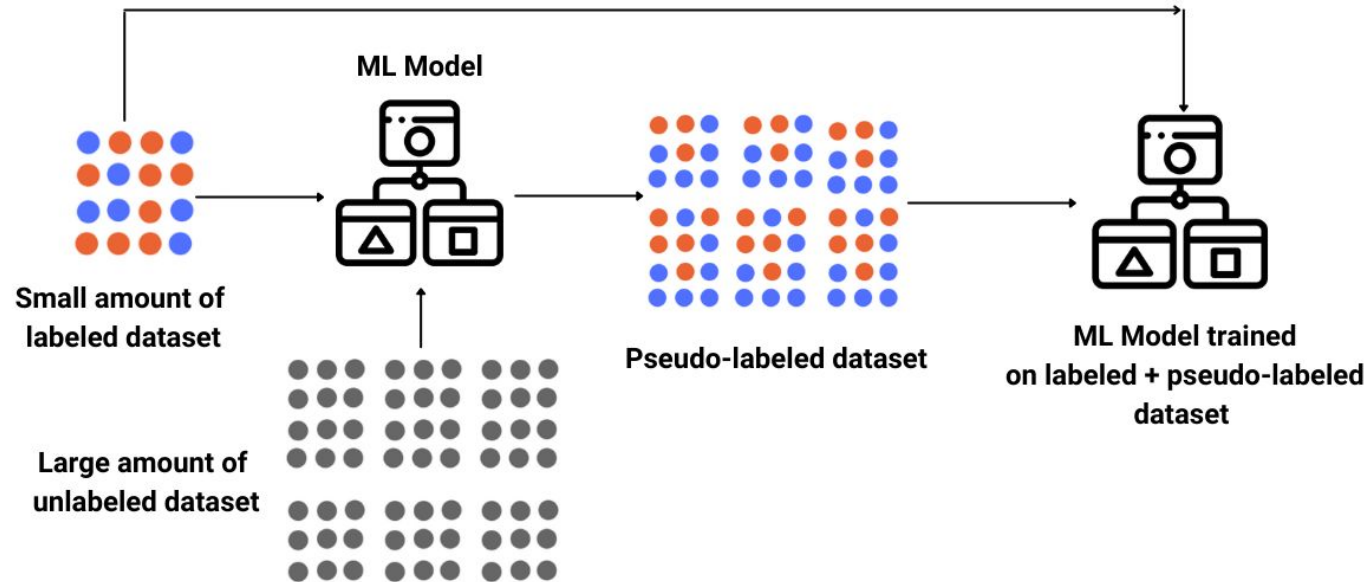


Fig. 1. General idea of Semi supervised learning
(<https://www.enjoyalgorithms.com/>)

Learning paradigms in SSL

Two dominant learning paradigms in SSL

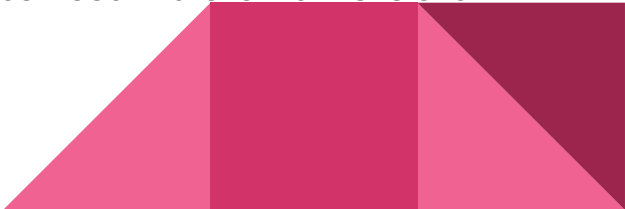
- **Transductive**
 - Transductive methods do not construct a classifier for the entire input space. So the predictions of such systems are restricted to the objects on which they have been trained and therefore, transductive methods have no separate train and test phases.
- **Inductive**
 - A classifier that relies on inductive methods can predict any object in input space. This classifier may be trained using unlabeled data, but its predictions for previously unseen objects are independent of each other once training is complete.

The majority of graph-based methods are transductive, whereas most other types of SSL methods are inductive.



SSL Assumptions

SSL may not be improved unless certain assumptions are made about the distribution of data.

- **Self-training assumption:** Predictions with high confidence are considered to be accurate.
 - **Co-training assumption:** Instance x has two conditionally independent views and each view is sufficient for a classification task.
 - **Generative model assumption:** When the number of mixed components, a prior $p(y)$, and a conditional distribution $p(x|y)$ are correct, data can be assumed to come from the mixed model.
 - **Cluster assumption:** Two points x_1 and x_2 in the same cluster should be categorized as one.
 - **Low-density separation:** A low-density region should be used as the boundary, not an area with high density.
 - **Manifold assumption:** If two points x_1 and x_2 are located in a local neighborhood in the low-dimensional manifold, they have similar class labels.
- 

Taxonomy of DSSL

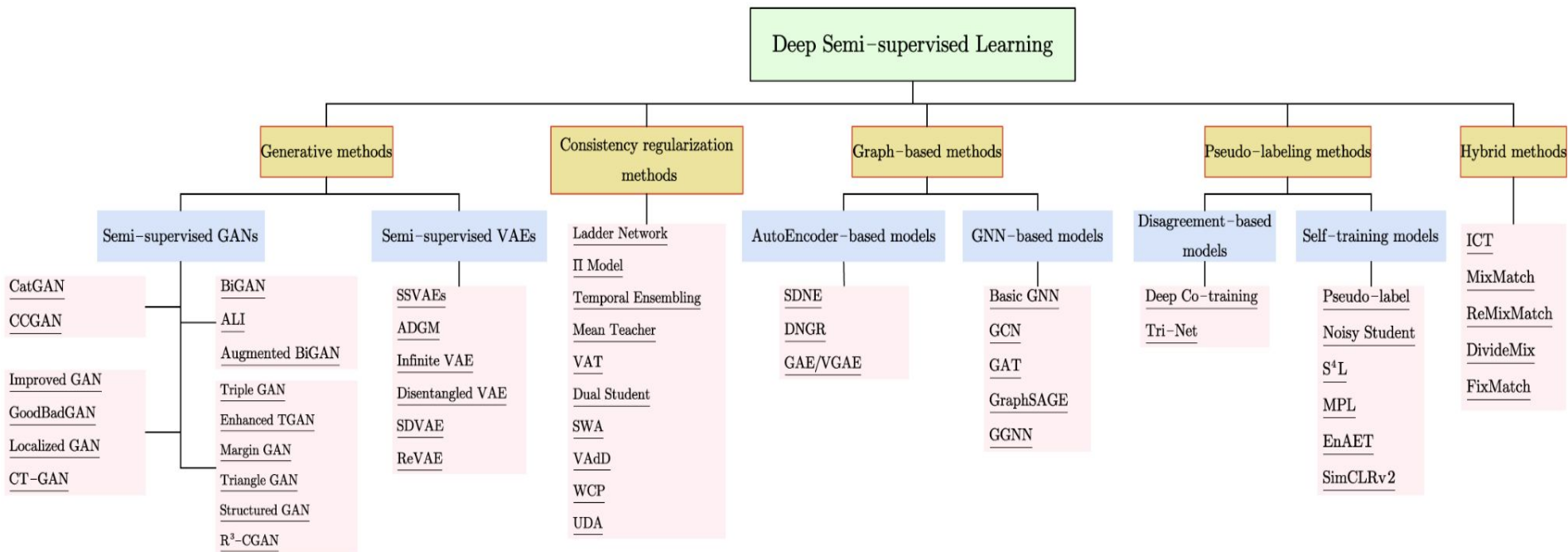


Fig. 2. The taxonomy of major deep semi-supervised learning methods based on loss function and model design.

Important DSSL methods

- Generative models
 - Semi-supervised GANs
 - Semi-supervised VAE
- Consistency regularization methods
- Graph based methods
 - Auto-encoder based methods
 - GNN based methods
- Pseudo labeling methods
 - Disagreement based models
 - Self training models
- Hybrid methods

Generative models - Semi-supervised GANs

- A GAN is an unsupervised model. It consists of a generative model that is trained on unlabeled data, as well as a discriminative classifier that determines the quality of the generator.
- GANs are able to learn the distribution of real data from unlabeled samples, which facilitates SSL. There are four main themes in how to use GANs for SSL.
 - re-using the features from the discriminator
 - using GAN-generated samples to regularize a classifier
 - learning an inference model
 - using samples produced by a GAN as additional training data.



Semi-supervised GANs - Examples

- **CatGAN:** Categorical Generative Adversarial Network (CatGAN) modifies the GAN's objective function to incorporate mutual information between observed samples and their predicted categorical distributions. The goal is to train a discriminator which distinguishes the samples into K categories by labeling y to each x, instead of learning a binary discriminator value function.
- **CCGAN:** The use of Context-Conditional Generative Adversarial Networks (CCGAN) is proposed as a method of harnessing unlabeled image data using an adversarial loss for cases like image in-painting. Contextual information provided by the surrounding parts of the image. The generator is trained to generate pixels within a missing piece of image.
- **Improved GAN:** There are several methods to adapt GANs into a semi-supervised classification scenario. Cat-GAN forces the discriminator to maximize the mutual information between examples and their predicted class distributions instead of training the discriminator to learn a binary classification.

Generative models - Semi-supervised VAE

Variational AutoEncoders (VAEs) combine deep autoencoders and generative latent-variable models.

VAE is trained with two objectives - reconstruction objective between inputs and reconstructed versions, and variational objective learning of a latent space that follows a Gaussian distribution.

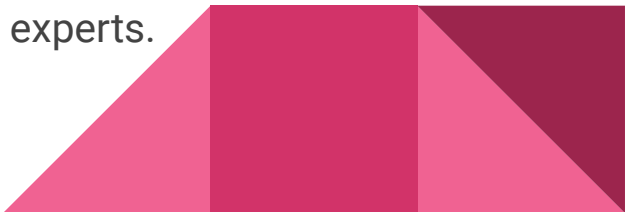
SSL can benefit from VAEs for three reasons:

- Incorporating unlabeled data is a natural process
- Using latent space, it is possible to disentangle representations easily.
- It also allows us to use variational neural methods.


VAEs can be used as semi-supervised learning models in two steps.

- First, a VAE is trained using both unlabeled and labeled data in order to extract latent representations.
- The second step entails a VAE in which the latent representation is supplemented by the label vector. The label vector contains the true labels for labeled data points and is used to construct additional latent variables for unlabelled data.

Semi-supervised VAE - Examples

- **Semi-supervised Sequential Variational Autoencoder (SSVAE):** Consists of a Seq2Seq structure and a sequential classifier. In the Seq2Seq structure, the input sequence is firstly encoded by a recurrent neural network and then decoded by another recurrent neural network conditioned on both latent variable and categorical label.
 - **Infinite VAE:** Mixture of an infinite number of autoencoders capable of scaling with data complexity to better capture its intrinsic structure. The unsupervised generative model is trained using unlabeled data, then this model can be combined with the available labeled data to train a discriminative model, which is also a mixture of experts.
- 

Consistency Regularization

- Consistency regularization is based on the idea that predictions should be less affected by extra perturbation imposed on the input samples.
 - Consistency regularization SSL methods typically use the Teacher-Student structure.
 - The objective is to train the model to predict consistently on an unlabeled example and its perturbed version.
 - The model learns as a student, and as a teacher, it creates targets simultaneously. Since models themselves generate targets, they may be incorrect and are then used as students for learning.
- 

Consistency Regularization - Examples

- **Ladder networks:** This network consists of two encoders, a corrupted and clean one, and a decoder. The corrupted one has Gaussian noise injected at each layer after batch normalization. The input x on passing through clean encoder will produce output y and on passing through corrupted one produce \tilde{y} . This \tilde{y} will be fed into the denoising decoder to reconstruct the clean output y . The training loss is computed as the MSE between clean activation and reconstructed activation.
- **Π -Model:** This is a simplified ladder network, where the corrupted encoder is removed, and the same network is used to get the prediction for both corrupted and uncorrupted input. In this model, two random augmentations of a sample for both labeled and unlabeled data are forward propagated, and the objective is to produce consistent predictions on the variants, ie reduce the distance between two predictions.

Graph-Based Methods

- The main idea of graph-based semi-supervised learning (GSSL) is to extract a graph out of the raw data where each node represents a training sample and the edge represents the similarity measurement of samples.
- There are labeled and unlabeled samples, the objective is to propagate the labels from the labeled nodes to the unlabeled ones.
- The GSSL methods are broadly classified into two:
 - AutoEncoder-based
 - GNN-based methods.



Graph-Based Methods - Examples

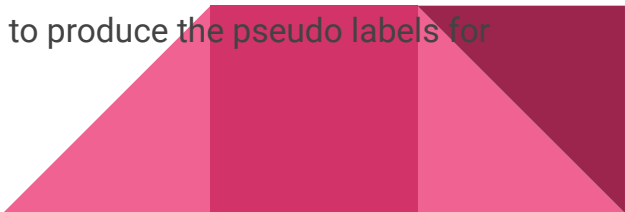
- **Structural deep network embedding (SDNE)** is an AutoEncoder based method. This framework consists of unsupervised and supervised parts. The first is an autoencoder designed to produce an embedding result for each node to rebuild the neighborhood. The second part utilizes Laplacian Eigenmaps, which penalize the model when related vertices are far apart.
- **Basic GNN:** Graph Neural Networks (GNNs) is a classifier is first trained to predict class labels for the labeled nodes. Then it can be applied to the unlabeled nodes based on the final, hidden state of the GNN-based model. It takes advantage of neural message passing in which messages are exchanged and updated between each pair of nodes by using neural networks.

Pseudo-Labeling Methods

The pseudo labeling method works in two steps.

- In the first step, a model is trained on a limited set of labeled data.
- The second step leverages the same model to create pseudo labels on unlabeled data and add the high confidence pseudo labels as targets to the existing labeled dataset creating additional training data.

There are two main patterns: one is to improve the performance of the whole framework based on the disagreement of views or multiple networks, and the other is self-training.

- Disagreement-based methods train multiple learners and focus on exploiting the disagreement during the training.
 - The self-training algorithm leverages the model's own confident predictions to produce the pseudo labels for unlabeled data.
- 

Pseudo-Labeling Methods - Examples

- **Pseudo-label:** This is a simple and efficient SSL method that allows the network to be trained simultaneously with labeled and unlabeled data. Initially, the model is trained with labeled data using a cross-entropy loss. The same model is used to predict an entire batch of unlabeled samples. The maximum confidence prediction is called a pseudo-label.
- **Noisy Student:** This is a semi supervised method that works on knowledge distillation with equal-or-larger student models. The teacher model is first trained on labeled images to generate pseudo labels for unlabeled examples. Following this, a larger student model is trained on the combination of labeled and pseudo-labeled samples. These combined instances are augmented using data augmentation techniques and model noise. With several iterations of this algorithm, the student model becomes the new teacher and relabels the unlabeled data, and the cycle is repeated.
- **SimCLRv2:** This is the SSL version of SimCLR (A Simple Framework for Contrastive Learning of Visual Representations). SimCLRv2 can be summarized in three steps: task agnostic unsupervised pre-training, supervised fine-tuning on labeled samples, and self-training or distillation with task-specific unlabeled examples. SimCLRv2 learns representations by maximizing the contrastive learning loss function.

Hybrid Methods

Hybrid methods combine ideas from the above-mentioned methods such as :

- Pseudo-label
- Consistency regularization
- Entropy minimization for performance improvement.

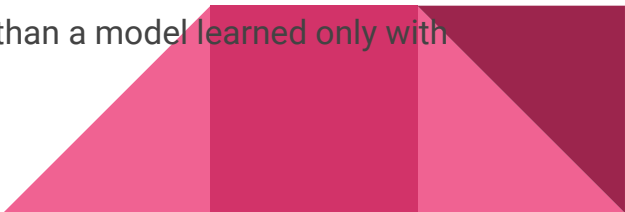


Hybrid Methods - Examples

- **MixMatch:** This method combines consistency regularization and entropy minimization in a unified loss function. It first introduces data augmentation in both labeled and unlabeled data. Each unlabeled sample is augmented K number of times and predictions of different augmentations are then averaged. To reduce entropy, guessed labels are sharpened before a final label is provided. After this, Mixup regularization is applied on both labeled and unlabeled data.
- **FixMatch:** This method combines consistency regularization and pseudo-labeling in a simplified way. Here, for each unlabeled image, weak augmentation and strong augmentations are applied to get two images. Both augmentations are passed through the model to get predictions. Then it uses consistency regularization as cross-entropy between one-hot pseudo-labels of weakly augmented images and prediction of strongly augmented images.

Challenges

The SSL method, like any other machine learning method, has its own set of challenges.

- One of the major challenges is that it is unknown how SSL works internally and what role various techniques, such as data augmentation, training methods, and loss functions, play exactly.
 - The SSL approaches above generally operate at their best only in ideal environments when the training dataset meets the design assumptions, however, in reality, the distribution of the dataset is unknown and does not necessarily meet these ideal conditions and might produce unexpected results.
 - If training data is highly imbalanced, the models tend to favor the majority class, and in some cases ignore the minority class entirely.
 - The use of unlabeled data may result in worse generalization performance than a model learned only with labeled data.
- 

Conclusion

Deep semi-supervised learning has already demonstrated remarkable results in various tasks and has garnered a lot of interest from the research community due to its important practical applications.



Thank you!