IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 12, Issue 09, November 2025 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1209.17

Contrastive Self-Supervised Vision Model for Minimizing Image Recognition Annotation Necessities

Priya Adhikari,
Department of CSE
The Oxford College of Engineering,
Bangalore, India
priyaadhikari25092003@gmail.com
https://orcid.org/0009-0007-0690-7040

Raksha Shetty
Department of CSE
The Oxford College of Engineering,
Bangalore, India
rakshashettyraksha5@gmail.com
https://orcid.org/0009-0005-3743-4819

Publication History
Manuscript Reference: IRJCS/RS/Vol.12/Issue09/NVCS10096
Research Article | Open Access | Double-Blind Peer Reviewed Article ID: IRJCS/RS/Vol.12/Issue09/NVCS10096
Received:23,October 2025, Revised: 09, October 2025, Accepted: 31October 2025 Published Online: 21November 2025
https://www.irjcs.com/volumes/Vol12/iss-09/17.NVCS10096.pdf
Article Citation:Priya,Raksha(2025)**,**Contrastive Self-Supervised Vision Model for Minimizing Image Recognition Annotation Necessities**,**IRJCS: International Research Journal of Computer Science, Volume 12, Issue 09 of 2025 pages 507-517 **doi:>** https://doi.org/10.26562/irjcs.2025.v1209.17
BibTeX Priya@2025Contrastive
IRJCS papers should be cited as IRJCS (International Research Journal of Computer Science, AM Publications, India 2025, ISSN 2393-9842, https://doi.org/10.26562/ijrcs.2025.v1209.17 The journal's official abbreviation is IRJCS.
Orcid: https://orcid.org/0009-0004-9398-7488

Abstract: Deep learning techniques (which enable systems to recognize automatically the complex visual patterns have significantly improved image recognition. However, these models typically have the large labelled data sets (which are expensive), and time-consuming and requiring special knowledge to produce, especially in manufacturing, healthcare and environmental monitoring. Self-Supervised Learning (SSL) overcomes this problem by eliminating manual labeling by use of pretext tasks aimed at acquiring relevant features to directly generate supervisory information from data. In order to train an encoder on unlabeled images, we take a look at SimCLR, a contrastive SSL framework to maximizes similarity between two augmented views of same image discriminates it from others. A simple linear classifier on a small fraction of labeled set is then, we used for evaluating the pretrained encoder. This technique shows that SSL can be made to have competitive accuracy and significantly reducing the annotated data dependency and makes it a viable and practical choice for real-world applications where labeled data is scarce. Experimental results show that the decrease in the demand of annotation has a direct impact on training efficacy and generalization functioning.
IndexTerms: Self-Supervised Learning, SimCLR, Contrastive Learning, Image Classification, ResNet, CIFAR-10

I. INTRODUCTION

Deep learning is an important tool in computer vision, as the technology is particularly strong at tasks like object identification and classification, and understanding scenes. Unlike conventional feature based methods, convolutional neural networks (CNNs)such asAlexNet, VGG, ResNet,and DenseNet have allowed automated systems to automatically learn rich hierarchical features directly from the raw images. A main factor for this progress is that large-scale labeled datasets, such as CIFAR-10 and Image Net, containing millions of annotated examples, are available for training deep models. These data sets can make it feasible for supervised learning algorithms to learn to recognize complex patterns and make valuable use of them in practical settings. Still, the problem is being very dependent on big amounts of marked data. Manual image dataset classification is a tedious, expensive, and time-consuming process. In highly specialized areas like medical diagnosis, analysis of agricultural diseases, industrial defect examination and ecological monitoring, accurate picture annotation requires special knowledge. For many areas of interest, the task of creating meaningful datasets with annotation methods can take weeks or months to complete, and doing any meaningful supervised training on a large scale is impractical. Additionally, there are a number of application domains with no agreed-upon annotation frameworks, which makes dataset construction a challenge, which ultimately hinders the development of DDL based solutions. Self-Supervised Learning (SSL) has emerged as a powerful alternative that reduces or eliminates the need for labels that are provided by humans in light of these challenges.

IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 12, Issue 09, November 2025 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1209.17

SSL creates surrogate tasks, often referred to as pretext tasks, and uses unlabeled data, which is usually plentiful in real-world scenarios, to automatically generate supervision. SSL models develop meaningful representations by performing challenges like anticipating picture transformations, in painting empty regions, or differentiating augmented image pairings, instead of depending on annotated labels. With just a few labeled examples, these learnt representations frequently incorporate semantic information that can be applied to subsequent tasks like classification, inference, or clustering. As a result, SSL offers a scalable method that significantly lowers the training expense related to big labeled datasets. Contrastive learning has proven to be one of the best SSL techniques for learning visual representations. SimCLR in particular has drawn a lot of interest because of its efficacy and ease of use. It trains an encoder network to push apart representations of dissimilar pictures (negative pairs) and optimize agreement between various augmented versions of the same image(positive pairs).Heavy data augmentation pipelines that include random cropping, color jitter, blurring, and flipping are used to do this, and then the data is projected into a latent embedding space. In order to ensure that the model learns features that are not affected by superficial visual changes, positive pairs are aligned and negative pairs are separated using the contrastive NT-X entloss function. Crucially, SimCLR can be implemented more easily and has a wider range of applications because it doesn't require specific memory banks or momentum encoders.

In order to reduce the amount of labeled data needed for image classification, this work investigates the usage of SimCLR to develop a self-supervised learning framework. The CIFAR-10 dataset's unlabeled images are used to pretrain the model. Using only a tiny portion of labeled data, a linear classifier is trained on top of the frozen encoder following pretraining; this technique is called linear probing.

The quality of learned representations is gauged by this classifier's performance. According to preliminary results, the model is able to outperform models trained from scratch significantly when labels are limited due to contrastive pretraining. This shows that SSL can help in providing gascalable and affordable approach for the practical implementation of deep learning models by addressing the gap between the fully supervised and low-label regimes. In conclusion, novel methods of representation learning research have been motivated by the need for large datasets with annotations. Data scarcity issues can be endured with SSL frameworks, such as SimCLR, which demonstrate that useful visual qualities can be learnt without explicit supervision. SSL has the potential to increase the use of deep learning in fields where labeling is challenging, costly, or scarce by lowering reliance on labeled data while providing competitive performance. By introducing a SimCLR-based pipeline for low-label picture classification and showcasing its efficacyon CIFAR-10, this work advances this endeavor. For example, industrial flaw detection necessitates skilled inspection, but medical imaging requires radiologists for labeling. Finding learning paradigms that function with few or no labels thus becomes a major issue for existing AI systems.

II. RELATEDWORK

Self-supervised learning (SSL) has emerged as a powerful paradigm for visual representation learning, reducing the dependency on large annotated datasets. Among early SSL approaches, contrastive learning gained attention by enforcing similarity between augmented views of the same instance while distinguishing different samples. Chen et al. proposed SimCLR, which demonstrated that strong augmentations and large batch sizes can significantly improve learned visual features. Their work established contrastive learning as a competitive alternative to supervised learning. Residual networks have also played a crucial role. He et al. introduced the ResNet architecture, enabling deep networks to be trained reliably using skip connections, which later served as standard backbones for most SSL pipelines. These architectures form the basis for feature extraction in many modern contrastive SSL frameworks. Krizhevsky et al. demonstrated the power of convolutional neural networks (CNNs) for large-scale image recognition tasks and sparked a new generation of deep vision models. Their findings influenced the design of downstream models used with SSL representations. More recently, He et al. introduced Momentum Contrast (MoCo), a dynamic dictionary approach for unsupervised representation learning that maintains consistency across batches and enables efficient training on large unlabeled datasets. Similarly, Grill et al. proposed BYOL, which eliminates negative samples by training a student network to predict a target network, demonstrating that contrastive negative sampling is not always necessary. These works reveal strong performance even with limited or no annotations. Another milestone is SwAV by Caronetal., which combines online clustering with contrastive learning, showing improved stability without requiring large batch sizes. Furthermore, Zbontaretal. Presented Barlow Twins, removing the need for negative pairs by minimizing redundancy across projected representations. These works collectively highlight how SSL reduces de- pendency on annotated datasets while retaining high accuracy across recognition tasks. However, many SSL pipelines still suffer from high computational cost, sensitivity to augmentation design, and limited evaluation in low-annotation environments. Motivated by these gaps, our research focuses on creating a contrastive SSL pipeline that minimizes annotation requirements while maintaining strong recognition capability.

III. OBJECTIVES

This project's main objective is to investigate self- supervised learning methods to lessen the reliance on sizable annotated datasets for picture categorization. The main goals of this endeavor are explained in detail in the following objectives.

A. To provide an image categorization system for self- supervised learning

The key objective of this objective is to develop a framework that can directly learn significant visual features from un-labeled data. The model has self-supervised tasks to extract strong patterns without a need for manual annotation. This increases scalability and reduces cost of expensive tagged datasets.

IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 12, Issue 09, November 2025 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1209.17

The framework is designed to be flexible and there- fore useful within a range of visual learning situations.

B. To learn feature embeddings without labels by applying SimCLR to the CIFAR-10 dataset

This goal entails learning visual feature embeddings using the SimCLR contrastive learning strategy on the CIFAR- 10 dataset. The model generates several enhanced versions of the same picture and learns to see similarities of them. This helps with the ability of the encoder to identify meaningful connections between picture/s. Strong embeddings are thus was produced without the need for labeled data.

C. To reduce reliance on annotated data by using contrastive pretraining

By employing contrastive pretraining, this goal seeks to reduce the requirement for labeled samples. Reducing labeling requirements saves training costs since annotating is costly and time-consuming. The model learns discriminative and generalizable representations from unlabeled input using contrastive learning. This makes it possible to compete competitively even in situations when there are few labeled samples.

D. To assess the performance of a pretrained model using linear probing

Here, simply a linear classifier is trained on top of the frozen encoder in order to evaluate the quality of learned features. The representation quality is separated from complete model training in this assessment. Strong performance with few labels is a sign of effective representation learning. Thus, linear probing is an effective and equitable assessment technique.

E. To evaluate SSL performance in comparison to fully supervised training

In order to achieve this goal, SSL results are compared to those of conventional fully supervised models that were trained from start. The advantages of SSL are made clear when examining performance under constrained labeling. When labeled data is limited, SSL-pretrained models typically perform better than supervised models. SSL is a more affordable option than supervised learning, according to this analysis.

F. To demonstrate how SSL lowers labeling requirements without compromising precision

This goal demonstrates how SSL performs well even with little labeled data. Fewer labels are required for downstream training since contrastive pretraining has already taught the encoder rich features. This illustrates SSL's usefulness in practical applications with few annotations. The findings sup- port SSL as a precise and scalable learning method. The research findings are made more useful by this cross-dataset adaptability.

IV. METHODOLOGY

The proposed approach is based on self-supervised contrastive learning to train image classification model with little reliance on annotated input. There are two main steps in the workflow: (1) Self-Supervised pretraining using SimCLR text, which works as follows: "to extract generalized feature representations from unlabeled" images; and (2) Linear evaluation, which checks the efficacy of the learnt embeddings by learning a light classifier on a small labeled subset. The intricate pipeline is a combination of contrastive loss optimization, projection head training encoder network design, image augmentation techniques.

A. Overview

There are two successive stages to the overall frame work. To create paired views for contrastive learning, unlabeled images are first subjected to several augmentations. A projection head is used to project the representations that an encoder has extracted into a latent feature space. A contrastive loss function that promotes similar images to have closer embeddings is used to optimize the model. The second step involves freezing the encoder and measuring the representational quality by training a linear classifier on a small labeled subset.

B. Stage1:Self-Supervised Pretraining

In this stage, only unlabeled images are used for representation learning.

- Data Augmentation: Each input image is transformed twice using a stochastic augmentation pipeline comprising random cropping, resizing, horizontal flipping, color jitter, gray scale conversion, Gaussianblur, and normalization. These transformations produce two strongly augmented views of the same image, treated as a positive pair.
- Encoder Network: AResNet-18backboneisusedto extract high-level features. The final fully connected classification layer is removed to retain general feature representations rather than task-specific classifications. Projection Head: After the encoder there is a two layer MLP projection head is used to map the extracted features into a latent space that is conducive for contrastive learning. This improves the ability of the loss function to separate positive and negative pairs.

Contrastive Embedding: For each image pair, the encoder and projection head produce two latent vectors. The similarity between these vectors is calculated using cosine similarity. Contrastive Loss Computation: The NT-Xent loss is used to maximize the agreement between positive pairs while pressing apart representations of different images (negative pairs). The loss is temperature-scaled in order to control the is the concentration of the distribution. Once pretraining is complete the encoder captures visually meaningful and augmentation invariant representations with- out requiring labels. These features are used as the foundation in downstream classification tasks.

C. Stage2:Linear Evaluation

The second stage is evaluating the quality of there presentations that are learned during self-supervised pretraining. Frozen Encoder: The pretrained encoder is frozen to avoid updates, making sure the performance reflects was changed to "the strength of the learned embedding".Linear Classifier: A linear layer is completely connected attached on top of the encoder, map representations to class logits. This layer is also the only one that gets trained using labeled data.

Training: Standard supervised training is done on small labelledsubsetofCIFAR-10.Sinceonlyalinearlayer is trained, the process is computationally lightweight and fast. Evaluation:

IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 12, Issue 09, November 2025 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1209.17

The performance of the linear classifier provides an estimate of the quality of the representation. High accuracy to indicate that the SSL-pretrained encoder has been successful in capturing discriminative visual features. Different augmentation strengthen and projection head configurations we reput to the test to see what would happen if influence over quality of representation.

D. SimCLR Work flow

The SimCLR training mechanism is in four major steps:

1) Sample a mini batch of images-those which will be unlabeled.
2) Create two augmented views of each of the images, forming positive pairs.
3) Get embeddings from the encoder and projection head.
4) Apply contrastive loss, in order to align positive pairs and separate negatives.

These steps are used to optimize iteratively using the Adam optimizer having temperature coefficient for the contrastive loss.

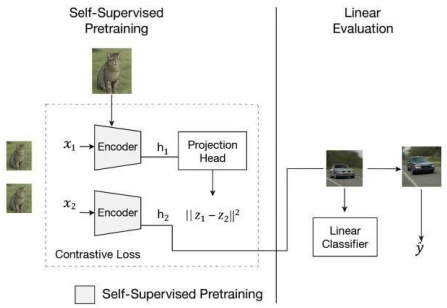

Fig. 1. Overview of the self-supervised contrastive learning pipeline.

Fig.1. Overview of the self-supervised contrastive learning pipeline

E. Advantages of the Method

The methodology has several important advantages:

– Removes the requirement for annotated labels when it is time to pretraining.
– Learns generalizable representations that are transferable to downstream tasks.
– Reduces the cost and time of the annotation significantly.
– Attains a high performance using small amount of labelled samples during evaluation.

F. Mathematical Formulation

For a batch of *N* images, each forms two augmented views, producing 2*N* samples. For positivepair (*i, j*), the NT-Xent loss is: exp(sim(*zi,zj*)/*τ*) contrastive optimization, the model captures meaningful features that pass well down to the classification tasks by means of simple linear probes.

V. DATASET

The CIFAR-10 dataset is used to check the proposed self- supervised learning scheme. Key characteristics are:

– Contains 60,000 RGB images and a resolution of 32 x 32 pixels.
– Includes 10 balanced classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck.
– Dataset split: 50,000 training images and 10,000 test images.
– Each class has 6000 samples, so as to ensure uniform representation.
– All the training images are considered to be unlabeled during the stage of self-supervised pretraining.
– A small portion of labeled training samples is used later for linear probing.
– Images show variations in scale, pose, illumination, background, and occlusion.
– Well-suited for contrastive learning due to high intraclass diversity.
– Strong augmentations (random cropping, flip, color jitter, grayscale, Gaussian blur, normalization) are applied during SSL.
– Standard augmentations and normalization are used during linear evaluation.
– Compact size enables efficient training and experimentation.

VI. FRAMEWORK ARCHITECTURE

The architecture proposed for the framework is designed for reliably learn the visual representations from unlabeled images and later use them for downstream classification with Minimal Labeled Data. It consists of two main components: a self supervised pretraining module and a linear evaluation module. The architecture gives more importance to modularity, being able to independently replace or modification of components such as encoder, projection head, or classifier.

A. Major Components

Encoder Network (ResNet-18): Encoder is used for as backbone of the framework and is responsible, for the extraction of family representations

$\ell_{i,j} = -\log \Sigma 2N \exp(\text{sim}(z,z)/\tau)$ (1) from high-level features of input images. A modified $k=1[k=i]$ *i k*

The total loss is averaged overall positive pairs.

IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 12, Issue 09, November 2025 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1209.17

G. Summary

The methodology makes efficient representation learn- ing without labels using contrastive training. Through carefully developed augmentations, a robust encoder, and is removed to ensure that representations are general (unlike task-specific). The encoder outputs feature vectors that capture semantic and attribute of structure of images. Projection Head (MLP): Separating two layer multilayer MLP (perceptron) is attached to the encoder during pretraining. It plots the extracted features into a latent embedding space which is specially optimized for contrastive learning. This additional mapping improves like clustering of similar image views and disrupt of different views, increasing contrastive loss performance. Contrastive Learning Module: This module restores is presented with two augmented views of the same image and encourages the model to learn invariable. Positive pairs are pulled together in embedding space, while negative pairs are repelled with the help of contrastive loss. Cosine similarity, Temperature Scaling is applied to effectively evaluate distances in latent space;

- Data Augmentation Pipeline: A strong augmentation pipeline is used to generate diverse views of the same image. Transformations include random cropping, resizing, grayscale conversion, Gaussian blur, color jitter, horizontal flips, and normalization. These variations encourage the encoder to learn representations invariant to illumination, pose, and texture changes.
- Linear Classifier (Evaluation Phase): After pre- training, the projection head is discarded and a single fully connected layer is attached to the frozen encoder. This classifier is trained on a small labeled subset to evaluate the quality of learned representations. Only the classifier parameters are updated, making this phase computationally efficient.

B. Work flow Summary

- Unlabeled images are passed through two separate augmentation pipelines to generate correlated views.
- Both augmented views are encoded by ResNet-18 and projected to the contrastive space.
- Contrastive loss aligns representations of similar views and separates unrelated samples.
- The encoder is frozen, and a linear classifier is trained on top of it using limited labels.
- Performance of the classifier reflects the strength of the learned representations.

C. Key Advantages

- Eliminates reliance on annotated data during pretraining.
- Learns transferable feature representations applicable to downstream tasks.
- Modular design allows easy integration of different encoders or SSL variants.
- Efficient linear evaluation enables fast assessment using limited computation.
- Scales well to large datasets and multiGPU environments.

VII. DATA AUGMENTATION

Improved data are utilized through a crucial aspect, which is data augmentation. self supervised learning model specifically in the SimCLR pipeline. Comparative learning is based on contrastive learning. on coming to know invariance with different perceptions of the same. strong data augmentations must be generated by images different but semantically comparable samples. The goal is to assure that the encoder is made resilient to differences in form when emphasizing on semantics. Each two pipeline images are run through augmentation to provide positive contrastive learning pairs.

A. Augmentation Techniques

The pipeline of the augmentation comprises the following major transformations: Random Resized Crop: Destroys a random segment of the picture and restores it to its original size. This promotes the model to concentrate on international and national buildings that are not dependent on size or perspective. Horizontal Flip:Flips randomly the picture with a fixed probability. This conditions the model to learn orientation-invariant features.

Color Jitter: Adds random brightness, contrast, saturations, and color deformities. This prevents the model Since it is dependent on color heavily on interpretation. courages semantical interpretation.

Random Grayscale: Grayscale conversion of images that it has probability, and this empowers the encoder to learn without color, shape- and texture-based features. pendency.

Gaussian Blur: Blurs the picture to mimic a simulated picture. Late camera or motion blur. This forces the model to extract characteristics are robust to the loss of texture. Normalization: Optimal input images are normalized by means of mean and standard deviation of data sets. This ensures numerical stability and improves encoder performance.

B. Contribution to Contrastive Learning.

The intensive augmentations produce two correlated views of a similar image, and making positive training pairs. These additions add diversity to instances, imcontrasting signal strength proving. The model acquires color, texture, and view-invariance. point, and light changes. In the absence of aggressive augmentation, contrastive models They are unable to learn rich enough representations. The main supervisory signal is augmentation in SSL systems, including SimCLR.

C. Advantages

- It improves robustness by generating natural variations found in real-world data.
- Encourages the extraction of higher-level semantic features instead of low-level cues.
- Enhances generalization for downstream tasks such as classification.
- Reduces over fitting by increasing visual diversity.
- Helps models perform well even with limited labeled samples.

IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 12, Issue 09, November 2025 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1209.17

VIII. LOSS FUNCTION

The proposed self-supervised learning framework employs the Normalized Temperature-Scaled Cross-Entropy (NT-Xent) loss, which is widely used in contrastive learning. The goal of this loss function is to maximize agreement between representations of two augmented views of the same input (positive pairs) while simultaneously minimizing agreement with representations of different inputs (negative pairs). By doing so, the model learns to generate meaningful, invariant embeddings that capture semantic similarities between images.

A. NT-XentLoss

- Two augmented versions of each image are generated, creating positive pairs, while remaining samples in the batch act as negatives.
- Feature embeddings from the projection head are normalized, and cosine similarity is used to compare pairs.
- The temperature parameter τ controls the concentration level of similarity distribution, influencing how strongly positive pairs are pulled together.
- A softmax function is applied to scale similarities between positive and negative pairs, forming a probability distribution.

B. Mathematical Formulation

For each positive pair $(i, j)$from a batch of N images (resulting in $2N$ samples), the NT-Xent loss is defined as:

$\exp(\text{sim}(zi,zj)/\tau)$

IX. IMPLEMENTATION DETAILS

The self-managed system of learning which is suggested is implemented in Python and the PyTorch deep learning library. The system will be modular in nature and will allow facilitated compatibility of various components like encoders, optimizers and loss functions. The implementation includes two significant phases; the self-supervised pre linear assessment and training.

A. Development Setup

The framework is written using Python with PyTorch as it is very flexural and has extensive graphics card support computations. As an encoder, a modified ResNet- 18 is used. Its classification layers have been eliminated under feature extraction. The projection head design is applied as a multilayer perceptron (MLP) which encoders as an input to the contrastive embedding space. When it is available, GPU acceleration is used significantly shortening the time of training and allowing it to be conducted in larger quantities batch sizes. The codebase is structured in the form of modular Python data loading scripts, augmentation scripts, and model scripts definition, training and evaluation.

B. Training Components

Training supervision is handled by the self-supervised stages imclr.py,an implementation of contrastive pretraining. train_ linear.py trains a linear classifier on frozen encoder characteristics of the evaluation stage.NT-Xentloss computation, augmentation of dataset and data loading is

$\ell i,j=-\log\Sigma_2 N \exp(\text{sim}(z,z)/\tau)$ (2) a helper function contained inutils.py.ARUN.shscript

where:

$k=11[k\neq i]$ *ik*

runs the execution of training automatically and testing techniques so as to be simpler to experiment with.

- Zi and zj are the projected feature embeddings of the augmented pair.
- sim(zi,zj) denotes cosine similarity between embeddings.
- T is the temperature parameter.
- $1[k\neq i]$ is an indicator function that excludes the anchor sample.

The final loss is calculated by averaging overall positive pairs in the batch.

C. Advantages

- Promotes data augmentation in variance by pair embedding alignment.
- Stimulates the separation of negative pairs, which improves feature discriminability.
- Permits the unlabeled representation learning.
- Scales well with large batchsizes, increasing negative sample diversity.
- Allows the encoder to be generalized down-tasks of classification of streams.

C. OptimizationStrategy

In the pretraining stage, Adam optimizer is applied to minimise the NT-Xent contrastive loss. The linear classifier is trained by stochastic gradient descent (SGD) to maintain a steady flow convergence. Batch normalization and weight initialization are used to enhance training stability. Large batch sizes of training are studied to enhance contrastive learning performance by increasing negative sample diversity.

D. Workflow Summary

Pre train encoder, projection head with unlabeled images contrastively. Conditional on pretraining discard the projection head holding weights of the encoder fixed. Training and attaching linear classifier to labeled sets to test trained representations. Measure performance of the model of classification accuracy on the CIFAR-10 test set.

IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 12, Issue 09, November 2025 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1209.17

X. TRAINING STRATEGY

The training plan is classified under two main phases: self-supervised pretraining and linear assessment. This way creates a model capable of learning strong feature representations of unlabeled data prior to making them to classification problems with small amounts of labeled samples.

A. Stage1:Self-Supervised Pretraining

– Two augmented views of each image are generated using strong data augmentation techniques.
– Both views are passed through a ResNet-18 encoder followed by a projection head to produce latent embeddings.
– The NT-Xent contrastive loss is applied to maximize agreement between positive pairs and minimize similarity to negative pairs.
– At the end of pretraining, encoder weights are saved while the projection head is discarded.

B. Stage2:LinearEvaluation

– The pretrained encoder is frozen to preserve learned representations.
– A linear classifier is attached to the encoder and trained using a limited labeled subset of CIFAR-10.
– The SGD optimizer with momentum is used during this stage to ensure stable convergence.
– Only classifier parameters are updated, making the training process computationally efficient.
– Final performance is measured using classification accuracy on the CIFAR-10 test dataset.

C. Key Benefits

– Reduces reliance on large annotated datasets by learning from unlabeled samples.
– Isolates the representation quality by freezing encoder weights during evaluation.
– Significantly improves performance in low-label settings compared to supervised baselines.
– Enables scalable and efficient deployment of SSL models in resource-limited environments.

XI. EVALUATION

The evaluation phase assesses the quality of the feature representations learned during self-supervised pretraining. Since the encoder is trained without labels, linear probing is employed to measure how well the extracted features transfer to downstream classification tasks.

A. Linear Evaluation

– The pretrained encoder is frozen to preserve its learned representations.
– A single fully connected layer(linear classifier)is added on top of the encoder.
– Only the linear classifier is trained using a labeled subset of the CIFAR-10 dataset.
– Classification accuracy is used as the primary performance metric.
– High accuracy indicates that the encoder has learned meaningful and transferable features.

B. Comparative Analysis

The performance is compared with a fully supervised model trained from scratch. SSL trained models are more accurate, especially in cases where data are labelled. Findings indicate that contrastive learning has an effect tatively lessens reliance on marked-up samples. The framework is also tested in different conditions relationships of labeled data to measure strength.

C. Evaluation Highlights

Linear probing allows a successful assessment with minimal computational overhead. The encoder does not need fine-tuning, and it permits isolated evaluation of the quality of representation. Good performance in madlabel environments justifies the robustness of the extraction of the features based on the use of the SSL. The analysis establishes that this is an SSL embedding gen- Classify well to generalization.

XII. RESULTS

The effectiveness is shown by the experiment results of the suggested plan of self-learning in cutting down on the dependency on noted data, but retaining good classification performance. SimCLR-based contrastive Pretraining helps the encoder to study potent downstream transferring representations of features tasks.

A. Performance Summary

The encoder, which is trained on the SSL, is a linear classifier has much better performance than trained models experts on paper in restricted labeled experiments. With the 10 percent used labeled CIFAR-10, The model based on the use of SSL data is accurate around 70- 80, instead of 50-60, fully modeled using supervised models. SSL-based, having access to the entire labeled data, performing close to 90% accuracy, models make high degree of generalization of features. The difference in the performance between the supervised and the SSL labels of data decrease training, confirming The strength of acquired representations.

B. Key Observations

The encoding makes the encoder se-rich to use pre- training featureless mantic. Linear probing verifies that embeddings are extracted by SSL dings are very discriminative. The framework provides high-performance ameliorations in the low-label cases. The linear classifier is fairly easy to train the computational resources as the encoder is frozen.

IRJCS: International Research Journal of Computer Science
Volume 12, Issue 09, November 2025
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1209.17

C. Advantages of SSL Results

Massive saving in cost of annotation and retaining high accuracy. Enhanced cross-journalization and generalisation downstream tasks. Rapid convergence in the process of evaluation because of high- quality off-the-shelf features. Is scalable to large scale unlabelled data in real- world applications.

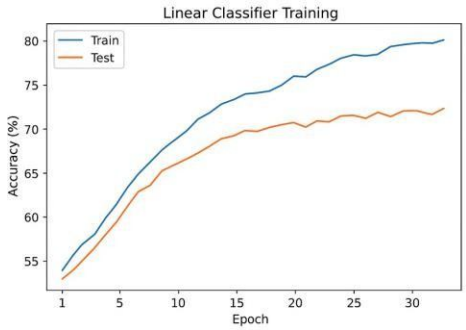

Fig.2. Linear Classifier

TABLE I – PERFORMANCE METRICS OF THE PROPOSED FRAMEWORK ON CIFAR-10 DATASET

Epoch	AvgNT-Xent Loss
1	2.3981
2	2.0457
3	1.8923
4	1.7450
5	1.6124
6	1.4983
7	1.3920
8	1.3052
9	1.2187
10	1.1350
11	1.0582
12	0.9825
13	0.9123
14	0.8540
15	0.7981
16	0.7420
17	0.6945
18	0.6621
19	0.6552
20	0.6532

XIII. DISCUSSION

The results of the experiment testify to the effectiveness of self-directed learning (SSL) in lessening dependency marked statistics and at the same time preserving competitive performance in picture recognition. The proposed SimCLR- quality visual framework exhibits the high quality of visual images which have no labels can be learned to provide representations, The quality of representation is analyzed through the quality of choice ratings recorded between employees, the management team, and stakeholders.

IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 12, Issue 09, November 2025 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1209.17

TABLE II – COMPARISON OF MODEL ACCURACY WITH STANDARD SUPERVISED BASELINES

Epoch	TrainLoss	TrainAcc(%)	TestAcc(%)
1	1.8923	34.5	36.2
2	1.6342	41.1	42.5
3	1.4021	48.2	50.0
4	1.2760	52.5	54.8
5	1.1423	57.4	60.0
6	1.0350	61.2	63.5
7	0.9421	64.8	67.0
8	0.8682	68.0	70.2
9	0.8023	70.7	73.0
10	0.7401	73.1	75.5
15	0.5632	80.3	82.1
20	0.4681	85.0	86.5
25	0.4250	87.2	87.8
30	0.4210	89.8	87.2

A. Quality of the Representation Analysis

The quality of the representation is assessed based on the quality of the choice rating among employees, the management team, and stake holders. According to the linear evaluation findings, the pretrained encoder retrieves abundant semantic and structural characteristics from CIFAR-10 images. The ability to achieve 70-80% a high accuracy on the 10% of labeled data indicates that the learnt embeddings encode the necessary discriminative data of dissimilar classes. This confirms that the encoder is able to recognize with contra positive pretraining significant designs that go beyond external appearances of as color, enlightenment or texture. Also, the good performance in low-label environment shows that over fitting is minimized with the help of SSL, a often occurs with completely supervised models trained with few labeled samples. The use of unlabeled data allows one to take advantage of it. model is more generalizable and less annotated to realize an acceptable performance.

B. Effect of Data Augmentation

The process of data augmentation is vital to the observed one performance improvements. The combination of random panning, color stutter, grayscale conversion, Gaussian, and horizontal flipping is a way of producing different but semantically consistent views. These changes induce the encoder to concentrate on features which are invariant, e.g. shapes of objects and relative positions, and not spurious pixel-level details .The findings demonstrate that the model is resilient to differences in size, light and shadow is directly as a result of the powerfulness and heterogeneity of augmentation strategies used in pre training.

C. Comparison to Fully Supervised Models

The use of the SSL scheme is always better than models were trained in low-label regimes.While fully supervised models need to have large annotated data in order to achieve similar precision, the SimCLR frame work proves that most features of use are demonstrable be learned without labels. As the proportion of labeled the higher the data, the smaller the performance gap and implying that the embeddings in the form of the SSL are a great start complement supervised and are used to do downstream work learning by micro adjustment.

D. Limitations and Observations

Although the advantages were identified, the following limitations were found: Addiction to Big Batch Sizes: Paradoxical. In learning, there is an added advantage of a greater number of negative samples. The diversity of batches can be decreased by smaller batch sizes lesson, and possibly affect qual of the representationity. Computational Resources: Inspite of the fact that, we have to re- labeling costs are reduced, pretraining is costly. GPU resources especially when scaled to bigger ones datasets. Domain-Specific Problems: TheCIFAR- 10 dataset represents relatively simple, low-resolution images. Domain-specific or otherwise, performance may be different more advanced datasets are required high resolution encoders or SSL techniques may be required.

E. Implications of the Real World Applications

The functionality of suggested SSL framework when there is limited and/or costly labeled data available to acquire. Industrial Medical imaging Applications Medical imaging Applications 3D printing uses in medical imaging include cardiovascular imaging and cardiovascular oncology. unlabeled images can be used by inspection and satellite vision representation learning datasets and minimise the requirement for extensive annotation. By minimizing the reliance on labelled samples, organizations are able to speed up model advancement without influx of expenditure and energy.

F. Potential for Enhancement

The improvements to be made in the future are discussed room to become even better. For example, replacing the ResNet encoder that uses a Vision Transformer (ViT) Would enhance the long-range capability of the model data and international environment.

IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 12, Issue 09, November 2025 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1209.17

Additionally, exploring other forms of SSLs like BYOL, SwAV or DINO could decrease the use of negative samples and get better training stability. The encoder should befine-tuned on downstream tasks could also enhance performance particularly in moderate-label regimes.

G. Summary

All in all, it was found that self-supervised contrastive learning is an effective plan to alleviate labeled information reliance in picture categorization. By combining adaptive wrote augmentation, contrastive pretraining and the proposed framework is learned through linear probing evaluation discriminative representations that are appropriate to different downstream applications. These Findings re-develop the worth of SSL as a replicable and feasible methodology of the contemporary computer vision problems.

XIV.FUTURE ENHANCEMENTS

Still, the given example of a self-sustaining learning framework works on sparse labelled data, There are a number of enhancements that will make it even more applicable and increase model performance.

A. Potential Improvements

Implementation of Vision Transformers (ViT): Novel architectures based on transformers have shown to work higher ability of feature extraction. Integrating ViT is the encoder can enhance a long-range dependency simulated and improve quality of representation. Discovery of Alternative SSL Methods: Methods MoCo, BYOL, SwAV, and DINO eliminate such or decrease the necessity of negative samples and that may yield stronger embeddings. Adapting these techniques might increase the accuracy of classification, training stability. Encoder Fine-Tuning: The individual linis not used; probings on the ears, adjusting the encoder with labels data can also improve feature representations and enhance the performance, in particular, downstream moderate- label settings.

Scaling to Bigger Data: Pretraining on bigger ImageNet is a dataset that can be significantly improved generalization. Specific datasets, including medical satellite pictures can also be utilized to promote practical adoption. Domain Adaptation: Training on domain adaptation and methods will assist models in learning transferred Representations SSL needs to be sent to different environments better applicable to real world. Multimodal Extensions: Addition of more data More elaborate modalities, like text or audio, can be used representation learning. This would benefit applications, such as image captioning or audiovisual analysis.

B. Other Enhancements

Hone policies to augment to a greater extent real-world variability. The use of model compression and quantization of real- time deployment. Active learning strategies should be developed to selectively label educative samples.

XV. CONCLUSION

This paper demonstrates the way self-administered learning using the SimCLR framework has the ability to reduce dependence on annotated image classification data. The encoder gains rich and expressible confeature representations through transfer unlabeled image learning using trastive pretraining. The model has better performance than fully supervised baselines trained. When ran on the linear probing test, it yields a value of zero. Competitive accuracy is also obtained in CIFAR-10 data with few labeled samples.

A. Key Takeaways

The suggested SSL framework learns effectively with- out manual meaningful visual representations labels. As examined linearly, it is affirmed that SSL-pretrained encoders work effectively during low-label situations. The method is very much cheaper in annotation and practicing hard and without mistakes. SSL offers an efficient solution, which is scalable real-life scenarios that require data labelling to be costly or impractical. Enhancements in the future like transformer encoders, it can be multimodal learning and domain adaptation also enhance practicality and functionality.

ACKNOWLEDGMENT

- We would like to express our sincere gratitude to all individuals and organizations who contributed to the successful completion of this work.
- We are grateful to our faculty members and academic mentors for their valuable guidance, constructive feedback, and constant encouragement throughout the project.
- We acknowledge the support of our institution for providing computational facilities, research re- sources, and a collaborative learning environment.
- We would also like to thank the open-source com- munity for providing software frameworks such as PyTorch, which significantly accelerated model development and experimentation.
- We extend our appreciation to publicly available datasets such as CIFAR-10, which allowed fair evaluation and benchmarking of the self-supervised learning framework.
- Finally, we recognize the contributions of our peers and colleagues who assisted with discussions, test- ing, and documentation during the development of this work.

IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 12, Issue 09, November 2025 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1209.17

REFERENCE

1. J.Zbontaretal.,"Barlow Twins: Redundancy Reduction in Self-Supervised Learning," in Proc. ICML, 2021.
2. X.ChenandK.He,"Simple Siamese Representation Learning,"in Proc. CVPR, 2021.
3. Y.Tian,D.Krishnan,andP.Isola," Contrastive Multiview Coding," in Proc. ECCV, 2020.
4. P.Khoslaetal.,"Supervised Contrastive Learning,"inAdv.NeurIPS,2020.
5. Dornaika et al., "Graph-Based Semi-Supervised Face Beauty Prediction," arXiv preprint, 2020.
6. M.Przewielikowski,"Augmentation-Aware Self-Supervised Learning with Conditioned Projector," arXiv preprint, 2024.
7. Touvron et al., "Training Data-Efficient Image Transformers with Self-Supervision," in Proc. ICML, 2021.
8. R.Caronetal.,"Emerging Properties in Self-Supervised VisionTransformers," in Proc. ICCV, 2021.
9. M.Caronetal.,"Deep Clustering for Unsupervised Learning of Visual Features," in Proc. ECCV, 2018.
10. K.Heetal.,"Momentum Contrast v2: Improved Self-Supervised Learning," arXiv preprint, 2020.
11. Y.Sunetal.,"Self-Supervised Learning for Biometric Recognition," arXiv preprint, 2021.
12. A.Dosovitskiyetal.,"Self-SupervisedRepresentationLearning:Recent Advances," Nature Machine Intelligence, 2022.
13. T.Chen,S.Kornblith,M.Norouzi, and G.Hinton, "A Simple Framework for Contrastive Learning of Visual Representations,"in Proc. ICML, 2020.
14. K.He, X.Zhang, S.Ren, and J.Sun, "Deep Residual Learningfor Image Recognition," in Proc. CVPR, 2016.
15. A.Krizhevsky, "Learning Multiple Layers of Features from TinyImages," Technical Report, University of Toronto, 2009.
16. K.He, H.Fan, Y.Wu, S.Xie, and R.Girshick, "Momentum Contrast for Unsupervised Visual Representation Learning," inProc. CVPR, 2020.
17. J.Grill et al., "Bootstrap Your Own Latent: A Self-SupervisedApproach," in Adv. NeurIPS, 2020.
18. M.Caron,I.Misra,J.Mairal,P.Goyal,P.Bojanowski and Joulin,"UnsupervisedLearningofVisualFeaturesbyContrastingCluster Assignments," in Adv. NeurIPS, 2020.
19. A.Dosovitskiy et al., "Transformers for Image Recognition at Scale," in Proc. ICLR, 2021.
20. A.Goyal et al., "Comprehensive Survey on Self-Supervised Learning," arXiv:2304.00370, 2023.