I. J. Engineering and Manufacturing, **2026, 4, 371-393**
Published Online on August 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijem.2026.04.25

Deep Convolutional Auto encoders with Structured Latent Space for Fashion-MNIST Denoising and Classification

Pattapu Sravani
Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
E-mail: pattapusravani2@gmail.com
ORCID iD: https://orcid.org/0009-0002-7284-351X

P. Satya srinivasa babu
Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
E-mail: satyasrinivas.p@kluniversity.in
ORCID iD: https://orcid.org/0000-0002-9597-4873

Inturu Bhavani Siva Phanindra
Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
E-mail: phaniinturu@gmail.com
ORCID iD: https://orcid.org/0009-0006-8957-7894

Shaik Hasane Ahammad
Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
E-mail: ahammadklu@gmail.com
ORCID iD: https://orcid.org/0000-0002-2587-4164

Ramachandran Thandaiah Prabu
Department of ECE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, SIMATS, Saveetha University, Chennai, Tamilnadu, India
E-mail: thandaiah@gmail.com
ORCID iD: https://orcid.org/0000-0002-5727-2977

Ahmed Nabih Zaki Rashed*
Electronics and Electrical Communications Engineering Department, Faculty of Electronic Engineering, Menouf 32951, Menoufia University, Egypt
E-mail: ahmed_733@yahoo.com
ORCID iD: https://orcid.org/0000-0002-5338-1623
*Corresponding Author

Received: April 3, 2026; Revised: April 22, 2026; Accepted: July 8, 2026; Published: August 8, 2026

Abstract: Fashion image analysis has a lot of trouble working with noisy data, especially when there aren't any good training examples to use. This happens because standard auto encoders can't keep class separability when the noise level changes. This paper introduces Structured Auto encoders (SAE), which combine convolutional encoder-decoder architectures with geometric constraints in the latent space. This strategy uses greedy layer-wise pre-training, and then it fine-tunes the model using two loss functions: structured latent space regularization and reconstruction error. The encoder has convolutional layers with 3×3 filters and ReLU functions that only turn on when they are needed. During testing, it was tested on Fashion-MNIST with noise levels ranging from 20% to 70%. It was also tested on MNIST, DeepFashion2, and 3D human pose datasets. The results of the experiment show that SAE can classify 85.4% of the time with only five samples labeled for each group. This is much better than the standard auto encoders (68.4%) and the baseline support vector machine (SVM) methods (84.32%). With the best setup, which has 1,000 hidden nodes and a learning rate of 0.1, images with up to 70% noise can be reconstructed well With 80% classification accuracy on clean test data, the latent space structured method allows for the solution of basic issues that are related to introducing geometric constraints between the classes so that the class boundaries remain intact even in a noisy environment.The

proposed method achieves a classification accuracy of 85.4% ± 1.2%under standard evaluation settings, with robustness validated under noise levels of up to 70%. All results are obtained using standard training–testing splits and are averaged over multiple experimental runs to ensure reliability and reproducibility. The proposed semi-supervised performance is enhanced by increasing the number of discriminative features through the use of a highly structured representation of the data, as opposed to using an unstructured representation, which results in performance enhancement. The authors present an original framework capable of simultaneously providing robust image classification and denoising capability through geometric constraints established in the framework, which enhances the learning of discriminative features through the use of limited amounts of labeled data. The implementation of this framework in the real-world fashion industry would facilitate the processing of fashion images with respect to noise resistance and the ability to annotate images quickly.

Index Terms: Convolutional auto encoders, structured latent space, fashion image analysis, denoising, semi-supervised learning

1. Introduction

Deep learning models, particularly autoencoders, have demonstrated strong capability in representation learning. However, their performance degrades significantly in real-world scenarios characterized by high noise levels and limited labeled data. Learning robust and discriminative latent representations under such conditions remains a challenging problem. The digital fashion industry has grown substantially, so we now have access to a larger quantity of visual data than ever. Processing this visual data in ways that will allow it to be meaningful for analysis and usage are necessary. Fashion image analysis employs all four methods of computer vision, deep learning, and pattern recognition. Fashion image analysis has the ability to automatically locate, organise, and group clothing items based on digital images. Auto encoders (which are a type of neural network) are becoming increasingly important for completing complex visual processing tasks, such as feature extraction, image reconstruction, and dimensionality reduction. They teach computers how to effectively and accurately represent visual data. Stability and reliability factors should be taken into account when designing auto encoder networks for effective virtual styling, automated fashion analytics, and eCommerce. This is important when dealing with fashion images because, due to differences in colour, texture, style, and background noise, it can be difficult to determine the subject of images. Auto encoders were developed as a result of modular learning. By utilising reconstruction-based criteria for feature extraction, these networks allow for the development of feature extraction without the need for supervision. Over time, deep architectures have proven to be better than traditional statistical methods at cutting down on the number of dimensions in data. This was possible because of pre-training strategies that made sure the network would start up correctly. These techniques were what made this happen. Adding convolutional layers to auto encoder designs made image processing much better. This was achieved by preserving the spatial structure [1] through the use of localized receptive fields. Structured latent spaces were added to probabilistic and adversarial frameworks to bring together the ideas of generative modeling and reconstructive learning. To enhance methods for noise removal [2, 3], noise was added to the inputs, and models were trained to restore clean representations. This made it much easier to find strong traits. The robust versions, on the other hand, tried to tell the difference between noise and outliers. They usually trained on clean datasets at first. They also tried to tell the difference between noise and outliers. To handle the greater complexity compared to digit recognition, customized datasets [2] were created for the fashion industry. The fashion industry made these datasets. These datasets contain a lot of pictures from many different categories. These pictures show how the texture, shape, pattern, and style have changed over time. Because the environment changes, automatic recognition systems have a hard time keeping their classifications accurate. Scientists have looked into neural ways to group things together and find small style differences. These methods are better than features made by hand, but they usually need a lot of data that is already labeled. To deal with the lack of annotations, semi-supervised learning has been used, and unsupervised pre-training has helped performance when there aren't many labels. Active learning strategies also try to improve labeling by concentrating on the samples that are most helpful. But there aren't many ways to combine structured representation learning with figuring out how sure something is. More and more research is now focused on structuring latent spaces so that the learned features are easier to understand and the important relationships between the features are kept. The primary focus of current structured latent spaces methods is on entities that provide better performance for representation learning/ clustering tasks, while the typical limitations include: limited use of only clean datasets and/or complete supervision to include the even more limiting factors of noise robustness and semi-supervisory learning combined create many barriers to real-life applications. Although much progress has been made by auto encoders in relation to the fashion image analysis and auto encoders application areas, many critical aspects are left unresolved. First of all, most auto encoder systems don't do a good job of keeping different classes separate in their internal representations, especially when they have to deal with images that are noisy or broken [4,5]. Most current denoising methods only care about making images look good again; they don't check to see if the cleaned-up images can still be

correctly classified [6]. Researchers are still trying to figure out how to make a system that can both accurately classify and reconstruct clear images at the same time. [7]. Another problem is that most of the methods we have now need a lot of labeled data to work well. This is a problem because it costs a lot of money and time to get fashion experts to label thousands of pictures eating [8, 9]. Lastly, we still need architectures that can handle different types and levels of noise across different fashion categories while still being practical enough to use in the real world [10]. We directly address these core problems by creating a new deep convolutional auto encoder that uses structured constraints in its latent space to improve and correctly classify fashion images at the same time. The main goal is to add different geometric limits. When noise makes images less clear, they help keep class boundaries clear and make it easier to tell features apart. In particular, we wanted to achieve the following: (1) build a structured auto encoder that keeps its internal representations of different classes separate, even when there is a lot of noise; (2) come up with an optimization method that uses dual loss functions to find the right balance between the quality of image reconstruction and classification accuracy; (3) test our method in semi-supervised settings with limited labeled data; and (4) show that our framework works with different types of fashion data. Our research looks at a lot of important areas. We're working on adding geometry-based constraints on auto encoders' latent spaces so that we can better classify between types of Fashion Images. We're also looking at optimising architectural changes to enable us to simultaneously optimise for both image reconstruction as well as classification in the presence of noise. In addition, we're considering the application of a structured latent representation when performing semi-supervised learning, where there are not enough Fashion Images with Accurate Labels. Finally, we will be comparing our proposed framework against different types of Auto-Encoder architectures and typical ML Algorithms as noise and complexity increase in the dataset. By using an experimental approach with convolutional auto encoder architectures with structured latent space regularization, we refined our methodology. We have two loss functions to find a balance between decreasing the number of reconstruction errors and maintaining the structure of the latent space. The Fashion-MNIST dataset is the primary dataset used for testing our method. However, we have also performed tests on the MNIST dataset, the DeepFashion2 dataset, and the 3D human pose dataset to show that our framework is general enough to apply to all these datasets. Our results are compared with those obtained from traditional auto encoders, variational auto encoders, and support vector machine benchmarks using an evaluation of how our method performs between 20% and 70% noise. The guided labeling method in our framework using the structured latent space provides an advantage by providing a mechanism to quickly choose samples to use in semi-supervised scenarios. Coupled with accurate confidence estimates from SAE and SVM, our method automatically selects instances with the highest uncertainty to gain the most benefit from hand-annotation.

To address these limitations, this paper proposes a structured autoencoder framework that integrates geometric constraints, noise-aware training, and uncertainty-driven guided labeling to enable robust learning under noisy and low-label conditions. Using classification scores for each class, the algorithm discovers samples that are very uncertain about the classification of that sample by calculating the highest score of the classes relative to the second highest classification for that sample. This method also reduces misclassification rates from about 4% to around 3%. The only manual task that needs to be performed with respect to the performance of the algorithm is to manually label the identified samples that were found to provide the most value to the algorithm. Therefore, it saves both time and money in terms of how much it would cost to document these samples manually. The framework is flexible enough to accommodate a wide variety of different types of data from vector data that describes 3D human poses with 6890 vertices, basic images that consist of MNIST images, to more complex images such as Fashion-MNIST images divided into summer, warm weather, and year-round clothing. Our method has demonstrated better performance than traditional classification networks in scenarios where there are very few training samples. The encoder-decoder design of our architecture has an advantage over many different network implementations (e.g., VGG-based implementations, fully connected dense networks [10, 11]). For example, when trained on only 6000 MNIST samples with some additional data, we attained a maximum accuracy of 99.05% on the MNIST test. Thus, the proposed approach will be beneficial for use in various applications of computer vision while maintaining the primary advantages of efficiency of directed annotation and structure of representation learning [12].

Structured latent space learning has been looked at in some of the previous literature. However, most of the current methods have been focused on providing a way to learn an abstract representation of a data set, and/or clustering the data without explicitly dealing with the impact of high noise and the limited amount of labelled data on their ability to represent the data. The framework that has been proposed provides a single structured autoencoder which uses the same geometrically constrained latent space for 3 tasks; denoising, classification, and semi-supervised learning. This work is unique because it includes four major innovations. The first innovation is how MDS-based geometric constraints were integrated with convolutional autoencoders to help enforce class separable structure from severe amounts of noise. The second innovation is the development of a dual-loss optimization approach that maintains both reconstruction fidelity and latent structure jointly. The third is an uncertainty-driven guided labeling mechanism to provide efficient semi-supervised training, and finally there is extensive validation on multiple datasets using different amounts of noise (up to 70% noise); both aspects have not been stress-tested in any other structured autoencoder frameworks.

The proposal makes the following contributions: 1. This paper proposes a novel structured autoencoder model comprising a convolutional architecture with a geometrically-restricted latent space to increase the separability of the classes; 2. The inclusion of a novel set of latent space constraints based on MDS will allow for the organisation of

features appropriately in the presence of significant noise; 3. A dual-objective loss function will jointly optimise for the accuracy of reconstruction as well as the structure of the latent space thus improving the ability to learn representations; 4. Provides an uncertainty-driven guided label mechanism will help to improve semi-supervised learning performance in the presence of limited labelled data; 5. The results of this research will be backed by extensive evaluations on multiple benchmarks, demonstrating that the novel approach achieves strong performance even with the extremely high level of noise introduced (up to 70%), which is not sufficiently addressed by existing structured autoencoder frameworks.

Subsequent sections layout how the subsequent parts of the study will proceed. The second part will lay out what we name the Proposed Structured Auto Encoder concept, how we develop the Architecture, and the way we set up our loss function, and ultimately how to properly Train/Test the proposed structured auto encoder with its respective parameters. The third section consists of the description of Experimental Methods (including the Dataset(s)) used for creating our results as well as the evaluation/measurement of those Results. The fourth section provides an analysis of each Auto Encoder using performance, ablation testing, and generalization analysis methods; also, the last part represents the specific usefulness of a specific Auto Encoder, or how a particular Auto Encoder type may be used. Additionally, the last half will provide a broad summary of the major findings and a potentially valuable future research area.

2. Related Work

In recent years, deep learning has undergone great advancements, particularly with respect to how it relates to representation learning via autoencoder-based models through the creation of structured latent spaces in which represented features can exist. The use of autoencoders to extract features, reduce dimensionality, and reconstruct data has led to many applications of these types of models; in fact, autoencoders form the basis of many unsupervised and semi-supervised learning frameworks. A recent survey, published in 2024, indicates that current trends in the design of autoencoders are moving towards using constraints on their latent spaces as a mechanism for increasing their interpretability and improving performance on downstream tasks [13].

Numerous studies conducted in 2024 have now emphasized developing latent space structures that promote improved feature separability and robustness through structural modifications of the underlying data during the generation of features. Latent space manipulation techniques enhance the quality of the representation to provide a better means for detecting anomalies [14] and models that utilize knowledge-integrated autoencoders impose distance-based constraints to help organize the latent space [14]. These efforts illustrate the advantages associated with the creation of structured representations but are primarily limited to a focus on representation learning rather than classification capability under conditions of noise.

The structured latent space autoencoders have been applied to domain-specific applications such as industrial monitoring and medical imaging in 2025. In particular, a structured deep autoencoder has been used in conjunction with reconstruction and latent representation learning for the purpose of condition monitoring [15, 16]. Other autoencoder-based generative models have also been used for image segmentation [17]. However, the majority of these techniques rely on large amounts of labelled data and therefore cannot effectively cope with semi-supervised learning environments where there is a lack of annotations.

Recent publications are exploring the improvement of latent space expressive power via more complex architecture ssuch as Graph Auto-Encoders and expanded latent space models, which allow improved feature learning and better classification performance [18], [19] but tend to lack a mechanism for accommodating high amounts of noise as well as maintaining class separation at the same time. In addition to this, most existing studies have examined either generative modeling or representation learning without considering an integrated approach that would satisfy both objectives within one unified framework [16]. In recent years, there have been major advancements in representation learning and semi-supervised learning, which have gained significant attention at leading conferences including CVPR, ICCV, and NeurIPS [20],[21],[22]. Many researchers have attempted to enhance model robustness and generalization ability using structured representations, contrastive learning, and noise-aware training strategies. All of these advances underscore the continuing value of learning discriminative latent spaces in difficult environments. Unfortunately, most of the current methods for accomplishing these tasks rely heavily on large amounts of labeled training data, or exclusively focus on either representation learning or classification independently (without combining the two into a single system that takes into account various geometric constraints and the need for semi-supervised approaches.

Most of the more recent work on structured autoencoders has focused on either increasing organization of latent spaces or improved learning representation; however, neither of these focused on providing a unique solution for both robustness against high-noise levels as well as providing a semi-supervised learning method when little or no data is available to use as ground truth for supervised training. Most existing methods assume a relatively clean dataset or need a majority of their training to be done in supervised mode. The proposed method therefore creates a framework that provides a unique combination of structured latent representation, noise-aware training, uncertainty-driven guided labeling and a single framework to accomplish this creates a very effective way of learning in an environment with

extreme input corruption, while providing very good classification performance with minimal supervision, therefore representing something which would be perceived as being entirely different than all existing approaches.

3. Proposed Methodology

3.1. Structured Auto encoder Framework

The overall structure includes various parts like: a convolutional autoencoder, geometric constraining of latent space through MDS, using SVD to align the data, classifying data with SVM classifiers and finally, adding an uncertainty-based active learning component. The training process occurs in the following sequence: first the autoencoder creates a latent representation of the data, then these representations have been geometrically constrained in order to give structure to the latent space, and finally these same representations have been aligned so that they can be used to train a classifier. The classifier will be iteratively improved using a small amount of labeled data by a sample selection process. Our method solves the main limitations of classic auto encoders by adding geometrical constraints into the latent space of a structured auto encoder (SAE) [23],[24] to obtain high fidelity (how accurately the original data can be reconstructed) and good classification (how accurately the data has been classified) of data. With explicit structural regularization providing us with a semantically organized learned feature space with separated classes. In contrast to traditionally learned auto encoders which only provide an unstructured latent representation of data, our innovation is the use of a dual optimization method that provides a balance between the accuracy of the reconstruction of the data and accuracy of the creation of the structured latent space [25]. The goal of this procedure allows us to both sort through and sanitize fashion-related imagery throughout this process while maintaining only the required components for subsequent tasks. Convolutional architectures are used to leverage the pre-existing spatial correlation found throughout each respective fashion image while utilizing 2D structural regularization via MDS-based approaches, to limit the actual geometry of the latent space. The arrangement of structural constrains within the latent space create limitations/restrictions to ensure represented data is optimally arranged representing similarity/interrelation of the data accordingly. The predefined structured Latent Space Z developed with the MDS optimization process assists in satisfying the distance preservation requirements for each represented class by enforcing the maximum inter-class distances while preserving the maximum intra-class distances throughout all class interrelationships. The latent space is set to a dimension of 64 for the autoencoder architecture, and the images that will be input into the autoencoder will be resized to 28 × 28 before being processed through the architecture itself. The structured latent representation, $\tilde{z}$, has been generated through a geometric transformation-based regularization process of the encoder's output, as outlined in (1).

$$\tilde{z} = R \cdot z \tag{1}$$

The rotation matrix R is derived from the SVD (singular value decomposition) of the structure matrix. As described in (2), R is calculated by: to align the learned latent representations with the desired target geometric structure, we adopted an orthogonal Procrustes formulation. The task is to find an orthogonal matrix R that minimizes the distance between the two matrices XXX and YYY.

$$\|XR-Y\|F\| XR - Y \|_F\|XR-Y\|F \tag{2}$$

The best solution comes from the method of computing the Singular Value Decomposition (SVD) using the cross-covariance matrix's SVD.

$$XTY=U\Sigma VTX^T Y = U \Sigma V^TXTY=U\Sigma VT \tag{3}$$

The rotation matrix is then given by:

$$R=UVTR = UV^TR=UVT \tag{4}$$

An orthogonal transformation that maintains distance and aids alignment with the target geometric form is achieved through this formulation. See (3) for the outline of the decoder function f_dec.

$$f\_dec(\tilde{z}) = \sigma(W_3^T * \sigma(W_2^T * \sigma(W_1^T * \tilde{z} + b_1^T) + b_2^T) + b_3^T) \tag{5}$$

Where the W_i are weight matrices and the b_i are the corresponding bias terms, a Decoder architecture is tied, i.e., the W_i of the Encoder is equal to W_i of the Decoder, to reduce the number of parameters to more manageable levels and increase performance with respect to generalization.

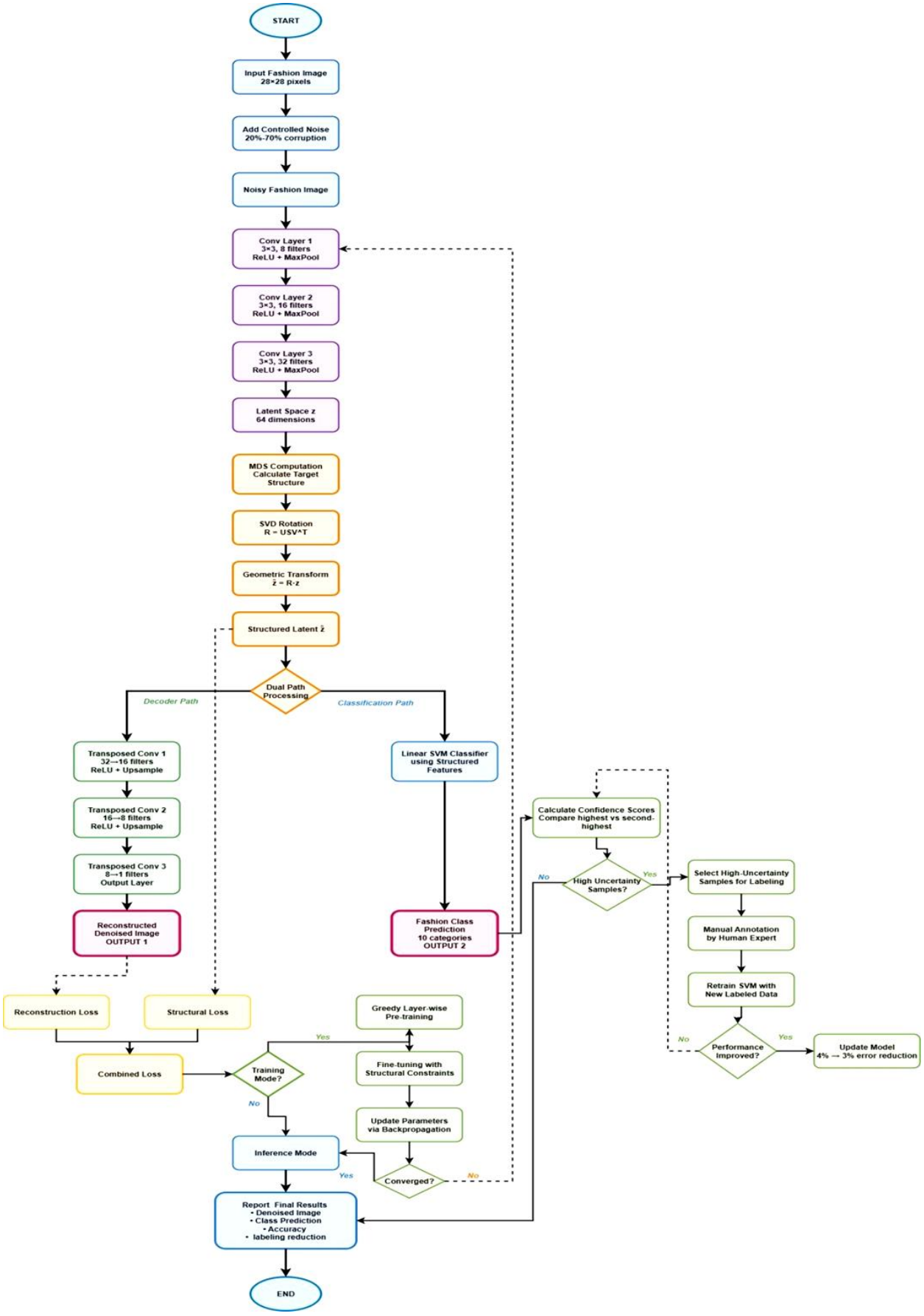

Fig. 1. Full Methodology for Structured Auto encoder with Guided Labeling Flowchart

As illustrated in Figure 1, the system has two routes that use Deep Learning to classify noisy images of Fashion items in a vigorous and efficient way. All fashion images start as 28x28 pixel images and have controlled Gaussian Noise of 20% to 70% added to them to simulate how the images would be damaged when captured from real-world video footage. To further facilitate these images for processing, we use ReLU activation functions and Max Pooling techniques to increase the size of each image by constructing a deep hierarchical structure. The framework contains three convolution layers which identify higher-level structures of shapes/patterns and textures. Additionally, implement Multidimensional Scaling (MDS) & Singular Value Decomposition (SVD) on the reconstructed images for an orthogonal rotation to improve how similar the structural elements appear in the latent representation. In order to preserve existing links between features, this method generates latent features that are aligned geometrically with each original dimension. Thus far, the structured latent space is considered to be the most important aspect within the model. This makes it easier to tell the parts apart and understand them. The framework then divides into two separate paths: one for decoding and rebuilding, and the other for sorting. Below, you can read about both of these methods. The decoder circuit can reconstruct the input while keeping the hidden features' spatial and contextual integrity by using transposed convolution layers. To sort the pictures into ten different fashion groups, a linear Support Vector Machine (SVM) is used in the classification process.

During training, the network optimizes a combined loss function that includes reconstruction loss, structural loss, and accuracy of classification. During training, this function is improved. This guarantees that discriminative learning and data integrity will always be in a state of balance with one another. First, a method called greedy layer-wise pre-training is used. After that, fine-tuning is done using structural constraints to make generalization and convergence more stable. There are many ways that an active learning loop can help the model work better. One of these methods is to use uncertainty sampling. To make the classifier work better, samples that are not very sure of their classification are found, labeled by professionals, and then added back to the training dataset, it will repeat this through the process of training a classifier. This has reduced classification errors by approximately 4% over the entire process of training multiple classifiers to classify items, making it a lot easier to classify an item. The classifier will now move into inference mode to analyze new data and provide the final estimates of performance in terms of classification accuracy, reconstruction-quality, and (if applicable) classification efficiency after convergence. The application of denoising, structural learning, and active training to this process produces a deep neural network that is able to achieve very high accuracy and produce interpretable features for fashion image classification in noisy environments.

In order to achieve structured class separation despite the introduction of noise, existing geometric constraints are applied in latent space. Through the use of multidimensional scaling (MDS), we ensure that distances between classes remain intact in the embedding of latent space. This provides a way for latent space to show meaningful geometric relationships between classes instead of just being a collection of learned features. To align the learned latent representation with the inherent geometric structure of the data, we also use singular value decomposition (SVD) based rotation. The orthogonal transformation maintains variance and preserves the pairwise distances between classes while improving the orientation of class clusters. Using MDS and SVD to construct a stable and interpretable mechanism to enforce global structure (such as demonstrated by MDS and SVD), will be especially useful in situations where supervision is weak and/or noise levels are high.

3.2. Loss Function Formulation:

The reconstruction loss measures if the image is built correctly and correctly applies the structured latent space principles; the reconstruction loss L_AE (the Mean Squared Error - MSE) can be defined as the amount of difference between the input image x and the reconstrcuted output f_ae(x) as in (6).

$$L\_AE(x, f\_ae(x)) = ||x - f\_ae(x)||^2 \tag{6}$$

The auto encoder learns through this part to capture useful visual representations that will preserve critical information that is important for accurately reconstructing the object or image. By using a squared L2 Norm to provide smooth gradients, the convergence will occur consistently throughout training. The structural loss part makes sure that the latent space stays organized by making sure that classes stay separate by following geometric rules. This is how the structural loss L_S is defined:

$$L\_S(f\_enc(x), \tilde{z}) = ||f\_enc(x) - \tilde{z}||^2 \tag{7}$$

This loss term makes sure that the encoder output matches the desired structured representation $\tilde{z}$. This limits the latent space to show useful geometric properties that improve classification performance and make the results easier to understand. The overall goal of the training is to combine reconstruction and structural loss into a single number by adding them together:

$$L\_SAE(x, \tilde{z}) = \gamma L\_S(f\_enc(x), \tilde{z}) + (1 - \gamma)L\_AE(x, f\_ae(x)) \tag{8}$$

Where, $\gamma \in [0, 1]$ is the balance parameter that controls how important structural goals are compared to reconstruction goals. The representation of the learned data structure has limitations based on the γ value. A scenario of $\gamma = 0$ will resemble standard auto encoders training. There is a trade-off when choosing a value for γ between the structural characteristics of the latent space and the fidelity when reconstructing the original inputs. Through empirical investigation, it has been shown that moderate levels of γ (generally between 0.3 and 0.7) produce an optimal balance between these two objectives, therefore allowing them to be easily optimized at the same time for classification and performance.

3.3. Training Procedures

Although not usually a downside since your program will still be able to run, having greedy layer-wise training of auto encoders prevents you from determining whether or not you've found the best solution to an instance of your choosing. However, performing layer-wise, greedy pre-training allows you to identify and establish good settings for your auto encoder's parameters (parameter initialisation). Pre-training also leads to excellent convergence properties when fine-tuning layers after pre-training layers. Layers are trained sequentially by feeding the outputs (i.e., weight values) of a preceding layer, into the corresponding neurons in a subsequent layer. This will gradually create more complex and deeper levels of feature representation from the first layer through deeper layers. The first layer (layer one) is trained using only the input pixels of images. The first auto encoder has learned how to represent simple features of the data, e.g., line (edges), surface (texture), orientation (simple patterns), in a manner similar to how the human brain processes visual information. Layer one will then pass its output (the evaluated weight values/layers) on to layer two as input; layer two will be able to determine more complex, relatable lateral connections using the layer one outputs rather than the raw input data. Layer two generally then forwards its output (obtain evaluated layer values) to layer three; thereby, the network can begin to capture increasingly complicated and hierarchical data patterns providing a rich set of features and depth for classification or reconstruction after completing auto encoding. As shown in (9), each layer training reduces the reconstruction error on its own.

$$\theta\_i^* = \text{argmin} \Sigma \, ||h\_\{i-1\} - f\_dec\^i(f\_enc\^i(h\_\{i-1\}))||^2 \tag{9}$$

Where, θ_i stands for the parameters of layer i, $h\_\{i-1\}$ stands for the output of the previous layer, and $f\_enc\^i$ and $f\_dec\^i$ stand for the encoder and decoder functions for layer i. After pre-training each layer, the whole network is fine-tuned with the combined loss function L_SAE. This stage adds structural limits while keeping the learned feature hierarchies from pre-training. Backpropagation is used in the fine-tuning process to optimize all of the network's parameters at the same time.

The fine-tuning optimization problem is set up like this:

$$\theta^* = \text{argmin} \Sigma \, L\_SAE(x\_i, \tilde{z}\_i) \tag{10}$$

where θ stands for all the network parameters and the sum goes over all the training samples. The structural constraints direct the optimization towards latent representations that demonstrate requisite geometric characteristics while preserving reconstruction abilities. Multi-Dimensional Scaling (MDS) is used to find the best class positions in latent space by calculating the geometric constraints. The goal of the MDS optimization is to keep the distance relationships set by a target distance matrix D.

$$Z^* = \text{argmin} \Sigma \, (d\_ij - ||z\_i - z\_j||_2)^2 \tag{11}$$

Where, d_ij is the distance that you want between samples i and j, and $||z\_i - z\_j||_2$ is the euclidean distance in latent space. The target distances are set to keep the classes separate from each other while keeping the classes close to each other.

We use Procrustes analysis to find the rotation matrix R in (11), which is used to line up encoder outputs with structured representations.

$$R = \text{argmin} \, ||Z\_enc - RZ\_target||^2\_F \tag{12}$$

Where, Z_enc stands for the encoder outputs, Z_target stands for the target structured positions, and $||\bullet||\_F$ stands for the Frobenius norm. As was said before, SVD decomposition gives us the best rotation matrix.

3.4. Structured Auto encoder Training Algorithm

Figure 2 shows the proposed structured auto encoder training algorithm [29]. It uses a systematic approach that starts with random parameter initialization and then trains each layer one at a time using reconstruction loss. We train each layer of the auto encoder separately and then freeze it to keep the learned representations. This keeps feature learning stable without interference from layers that haven't been trained yet. After the layer-wise pre-training is done,

the algorithm uses Multidimensional Scaling (MDS) on the desired distance matrix to find target structured positions. This imposes geometric constraints on how the latent space may be structured. The forward pass creates current latent representations, which are subsequently aligned with target locations using an ideal rotation matrix discovered using Singular Value Decomposition (SVD). Structured representations are created by imposing the required geometric structure while retaining crucial properties. Training is accomplished by repeatedly optimizing a balanced combined loss function, as illustrated in (11).

$$L = \gamma L\_S(z, \tilde{z}) + (1-\gamma)L\_AE(X, f\_ae(X)) \quad (13)$$

Where, the parameter γ controls the trade-off between the quality of the reconstruction and the alignment of the structure. In the backward pass, gradient descent is used to change the network parameters until the convergence requirements are met. This dual optimization method makes sure that the auto encoder can still reconstruct while learning structured representations that follow geometric rules. Optimized parameter values, θ*, representing structured auto encoders that have been completely trained, can be produced by this method for use in applications that require both high-quality reconstruction and interpretable organization of latent space. A systematic layer-by-layer approach to initializing training of network connections ensures that the stability of training dynamics is maintained and that the goals of both feature learning and structural constraints are achieved.

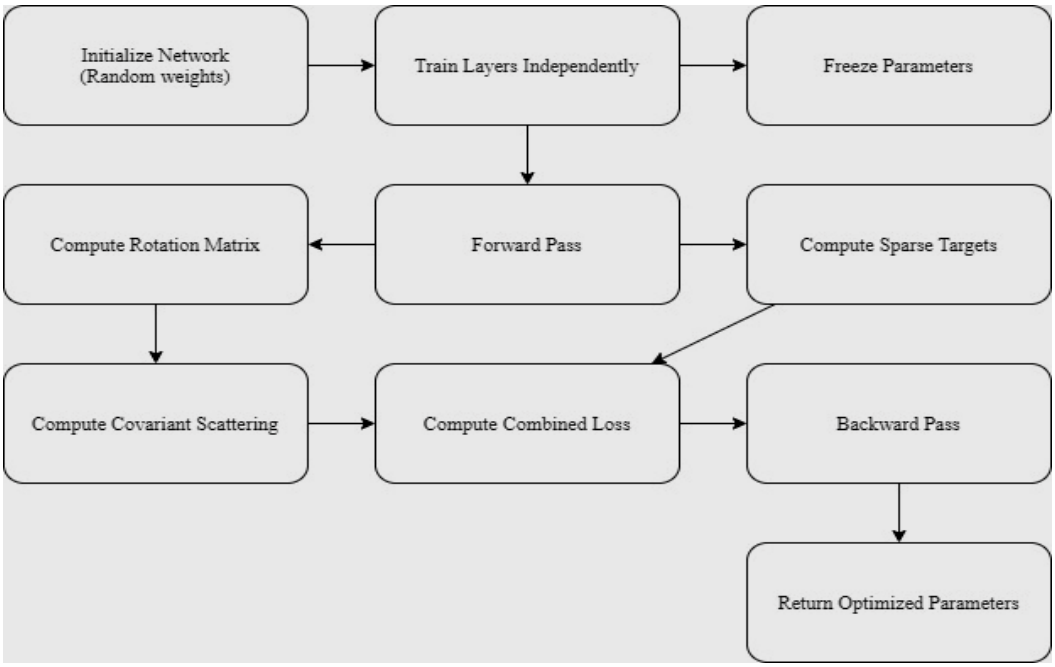

Fig. 2. Structured Auto encoder Training Algorithm [29].

In Fig. 2, the design of the latent spacegives assurance that the classes remain distinct from each other so that the trained auto encoder will be able to produce structured latent representations to be used as input features for classification by means of a linear support vector machine (SVM). The classification process consists of three steps. First, construct and train a classifier with training/explanatory variables from the structured feature and associated labels. Second, retrieve the latent representation $z\tilde{}$ = f_enc(x) with the trained auto encoder. Third, use the trained auto encoder-SVM framework to predict using new sample data. Using a two-phase approach to the semi-supervised learning of an SVM classification problem is facilitated by using the auto encoder, which was trained to extract features from unlabeled data, to improve classification of the target variable when trained with a limited number of labeled data points. The design of the latent space produces features that are component-based, so that the classification of sample data can be significantly enhanced regardless of the label provided during the classification phase. A linear SVM classification supported by a set of features extracted using a structured auto encoder is an effective and computationally efficient means of accomplishing problems where both the feature learning occurs without labels and when classification will occur with labels through supervision.

Algorithm 1: Proposed Training Pipeline

1. Initialize autoencoder parameters
2. Train autoencoder using input data to obtain latent representations

3. Apply MDS to enforce geometric structure on latent space
4. Perform SVD-based Procrustes alignment to refine latent representation
5. Train SVM classifier using available labeled data
6. Compute uncertainty scores for unlabeled samples
7. Select most informative samples (active learning)
8. Update labeled dataset
9. Retrain model with updated data
10. Repeat steps 5–9 until convergence

4. Experimental Setup

4.1. Dataset Description and Configuration

The proposed framework is evaluated on multiple datasets, including MNIST, Fashion-MNIST, Deep Fashion2, and a 3D pose dataset, to assess its robustness across varying data complexities. These datasets represent different domains, ranging from simple grayscale digit images to complex real-world visual data and structured pose representations. The use of multiple datasets is intended to demonstrate the adaptability and robustness of the proposed method under diverse conditions, rather than to claim direct cross-domain transfer learning. In this research study, we utilize the Fashion-MNIST dataset - which was released in 2017 and is considered a more difficult benchmark than the original MNIST - as the main dataset for our experimentation. Fashion-MNIST can better simulate real-world computer vision challenges with regards to fashion-related applications than the traditional MNIST dataset (handwritten digits) as it provides an increasing level of difficulty for both classification accuracy and time/efficiency, given the 70,000 gray scale images from multiple fashions (groups include T-shirt/top, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, and ankle boot) . The comparison and validation of Fashion-MNIST will be evaluated against the original MNIST dataset which contains 70,000 digit images (numbers 0-9) and are divided into three moderately-balanced subgroups of (0,1,9) (4,6,8) (2,3,5,7). There are 60,000 training samples in the Fashion-MNIST dataset, and 10,000 corresponding testing samples; both groups consist of 28x28 pixel (or 28/28 = 1.0) gray scale images with values ranging from 0-1. To create structured auto encoder analysis, images will be classified by the following seasonal categories: summer garments (T-shirt, sandal, garment and shirt), winter garments (pullover, coat and boot), and accessories for all seasons (trouser, sneaker and bag) . These categories will also provide input to the structured formation of latent space and create more meaningful geometric restrictions on learned representations. The dimensionality for Fashion-MNIST's latent space was reduced to 10 unsecured values (compared to 64 unsecured values in Fashion-MNIST) for less visually-complex representation of MNIST data samples. To show that the method can be used on more than just Fashion-MNIST, we validate a subset of the DeepFashion2 dataset that focuses on skirts and shorts that are on the edge of being too short or too long. It is hard to put these events into groups because they seem to be the same and have some things in common. The encoder network uses a convolutional implementation that is based on the VGG algorithm and has a latent space size of 192×192. This is done to keep up with the growing detail and complexity of real-world fashion photos. The SMPL is used to make a dataset of three-dimensional human poses so that different approaches can be compared across different types of data. There are 3,000 male and 3,000 female meshes in this collection. They were all made at random and come in a variety of body types and postures In total, each model contains 6890 vertices, each having x, y and z coordinates. The experiments were conducted using a consistent set of hyperparameters across all datasets. The input image size was fixed at 28 × 28, with a latent dimension of 64. The model was trained using a batch size of 128 and a learning rate of 0.001 for 50 epochs. The fully connected network architecture has two layers of thick layers (2048 and 256 neurons). This means that the vectorial data representation of the model is significantly different than the image-based representation of other images. It should be noted that the coordinate data are represented in vector format, which provides an alternative representation of the data. The latent space has 30 dimensions, and thus the three-dimensional pose data would provide a suitable means of representing the difficulty that humans experience changing positions as well as has an appropriate form for manipulation for any of the experimentations and Manipulation's designed for humans. All the experiments were run using standard dataset splits for training, validation, and testing. The performance metrics provided in this paper are computed by averaging several independent runs to guarantee statistical accuracy. The level of noise was added over a range of values to examine the robustness of the model under controlled situations. The same evaluation procedure is used across all datasets to provide consistency and fairness. All datasets are pre-processed uniformly and in a standardized manner. The pre-processing includes resizing and normalizing the datasets. The configuration of the model also remains the same for all experiments (i.e., the dimensions of the latent space, training parameters, and levels of noise injection). The evaluation for each experiment is conducted using the same evaluation procedure, and the performance of each experiment is computed as the average of several independent runs for statistical accuracy.

All experimental variables were held constant for all datasets and experiments unless stated otherwise. All methods were evaluated under the same experimental conditions (e.g., the same dataset splits, preprocessing steps, and evaluation metrics). Any differences between the supervision level and the amount of training data used, were documented and provided to allow for a fair and transparent comparison of methods.

4.2. Experimental Configuration

We wanted to assess how well our structured auto encoder could account for noise in images so we performed both systematic as well as random masking of input image files at various ratios (ranging from 20% to 70% corrupted) in order to see how corruption affected the performance of our model. In the real world, image corruption can occur due to a number of factors, but typically lies within a range of 20 - 70 percent corruption levels for a given image, with increments of ten percent. For each condition tested, we randomly selected a percentage of pixels to set to zero, similar to what we did in our situation. We have to deal with various forms and levels of corruption affecting an image in general, ranging from minor issues (20% corrupted) such as sensor noise, compression errors, etc. to significant issues (70% corrupted) such as a large area of occlusion, complete failure of transmission, etc. For all conditions tested, we repeated the same experiment multiple times for all levels, to ensure the validity of our statistical analysis. We also used identical code during both training and testing phases of our experiment. Our encoder consists of three convolutional layers, with three by three kernels, and varying numbers of additional filters (first layer with 8 filters, second with 16 filters, and third with 32 filters), all set to the same specifications. Our decoder, however, consists of transposed convolutions, using fewer filters than those used by the encoder (32 - 16 - 8 - 1), as well as tied weights ($W' = W^T$) in order to keep parameter numbers low. We set the final dimension of the encoder to the same as that of the input data (e. g., 256 x 256 = 65536, 65536 pixels in), while simultaneously ensuring that the dimension of the final layer of the decoder remains at 1 pixel. We evaluated the datasets' complexity and determined which reconstruction method produced the highest-quality reconstructions in the shortest time. We conducted extensive testing for hyper parameter tuning and found that the optimal learning rate for training was 0.1 when we tested various rates of 0.1, 0.2, 0.3, 0.4, and 0.5. The hidden layer size of 1000 nodes produced the highest quality of reconstruction. Our balancing value γ worked very well in the range of 0.3–0.7, which corresponds to a good trade-off between the structural loss of the original image and structural loss of the reconstructed image. We used 500 training iterations for our initial tests, and we trained our final models until they converged. We grouped our samples into sets of 5,000 images, allowing for an optimal balance of computation time and convergence stability. We used early stopping based on validation loss to prevent overfitting of the model. This also allowed us to obtain reliable statistics for benchmarking comparisons between the different datasets and different types of corrupted images.

4.3. Evaluation Metrics

We have evaluated the performance of our structured auto encoder on reconstructing objects from a set of different standards, with the primary method being Mean Squared Error (MSE). To measure the differences between the original and reconstructed images, we have used the Average of Squared Differences. We also used the Peak Signal-to-Noise Ratio (PSNR) on a decibel scale to measure quality in a way that is more like a logarithm. We also used the Structural Similarity Index (SSIM) to measure how similar things look because it is more in line with how people really see picture quality. In the same situation, both of these metrics were used. We used standard machine learning metrics to test how well our classification system worked. These were classification accuracy (which went up from 4% to 3% because of our guided labeling strategy), error reduction (which also went down from 4% to 3%), sample efficiency (to see how well we did with a small amount of labeled data), and annotation reduction (which tried to cut down on manual labeling by 100 times while keeping performance at the same level. Through these activities, we assessed how well guided labeling works by highlighting uncertainty sampling versus random selection when evaluating enhancement in performance over each cycle of active learning and by utilizing specific semi-supervised learning metrics to measure how well our designed latent space can help determine the most informative samples for our dataset. We compared our findings with other baseline models including traditional auto-encoders without structure, linear SVMs trained on unstructured pixel data, and probabilistic alternatives to auto-encoders such as VAEs (Variational Auto-encoders). In addition, we have also compared our own result to the current leading techniques from deep learning used for fashion image classification and denoising. Overall, we have shown our multi-metric comparison produced a comprehensive analysis of the reconstruction quality, classification rates, and evaluation of semi-supervised learning and additionally demonstrated that this methodology can be useful for uncertainty sampling under conditions where resources are limited.

4.4. Experimental Procedures

Validation involves performing a substantive amount of validation to determine whether any results are statistically significant and reliable. To create two (2) validation datasets from the original training dataset, one will use an approx. split of the original training dataset in half and will use a validation dataset to modify the hyper parameters of the model during training. The test set will not be modified. We used the 5-fold cross-validation procedure whenever possible to evaluate the performance of the model across various sample datasets and how consistent the results were with respect to those different datasets. Additionally, we used multiple random initialization seeds with each configuration because the training process is not always predictable, and to ensure that the resulting findings could be reproduced. We used paired t-tests during the verification of statistical significance to compare the performance of our proposed strategy with that of the baseline approaches. We can measure the practical value of the significant differences in performance using Cohen's d as a measure of effect size; this allows us to see how important all the benefits actually are when utilized practically. The use of systematic ablation studies ensures that each of the elements that contribute to

the overall strategy is constructed properly. There are three (3) components of the ablation study: structural constraint ablation ($\gamma = 0$ compared to $\gamma > 0$); directed labeling ablation (random versus predetermined/uncertainty -based); and architectural component analysis of distinct network topologies. Based on quantitative assessments of each component type, doing a full validation of all components (as used in this experiment) showed that every design element is necessary for optimal performance. A study of the amount of the controlling variable (γ) examined values from 0.0 to 1.0 in order to discover the best balance between structural and reconstructive objectives. It is common practice to establish 95% confidence intervals while evaluating key performance metrics; therefore, determining the amount of uncertainty associated with all possible outcomes is a valid method of measurement. All of the experiments were carefully designed, which helps to ensure repeatability of experimentation and reliability of the scientific results. Moreover, there are many articles available about training protocol, parameter tuning and setup procedure including many sources that offer assistance in the independent validation of machine learning methods. This represents multiple references that you will be able to consult during the independent validation process. The existing protocol is based upon directly comparing proposed and existing baseline methods on multiple performance measures and by obtaining extensive validation results of both methods from data sets that pass through the test framework. The goal of all this effort is to provide evidence that supports the conclusion that the proposed structure of the auto-encoder method does work and that the proposed auto encoder method meets acceptable levels of scientific validity, accuracy and reproducibility as measured by current standards for data collection and machine learning research. To demonstrate the effectiveness of this approach under different conditions, a systematic approach will be used, which utilizes statistical methods, ablation experiments, and an identical reproducible experimental design to enable the testing of all aspects of the systematic comparison and evaluation methods.

5. Results and Analysis

5.1. Hyper parameter Optimization Results

Finding optimal values for the learning rate parameters gives significant insight into the training of the structured auto encoder framework. The evaluation of learning rates in the range (0.1 - 0.5) all confirmed a best performing learning rate of 0.1 to result in the least amount of reconstruction cost regardless of the hidden layer configurations. Using 5,000 training examples from the Fashion-MNIST dataset with 20% noise corruption and 500 training samples showed that learning rates greater than 0.2 - 0.5 caused instability during training and less than optimal convergence.In the Fig. 3., we can see that learning rates affect how long it takes to train an artificial neural network. At first, lots of time is saved if fast learning rates are used; however, ultimately your trained neural networks will not work as well if you choose a fast learning rate because your final model will perform poorly and it may take an undefined amount of time for your trained neural net to converge to some point. Our experiments show that if we choose a learning rate of 0.1, we balance both speed and ultimate quality; our average training accuracy across the different experimental setups was 71.96%, while the average test accuracy was 72%. However, using faster learning rates of 0.4 and 0.5 resulted in only test accuracies of 60.04% and 65.30%, respectively, which is quite a drop in terms of performance.

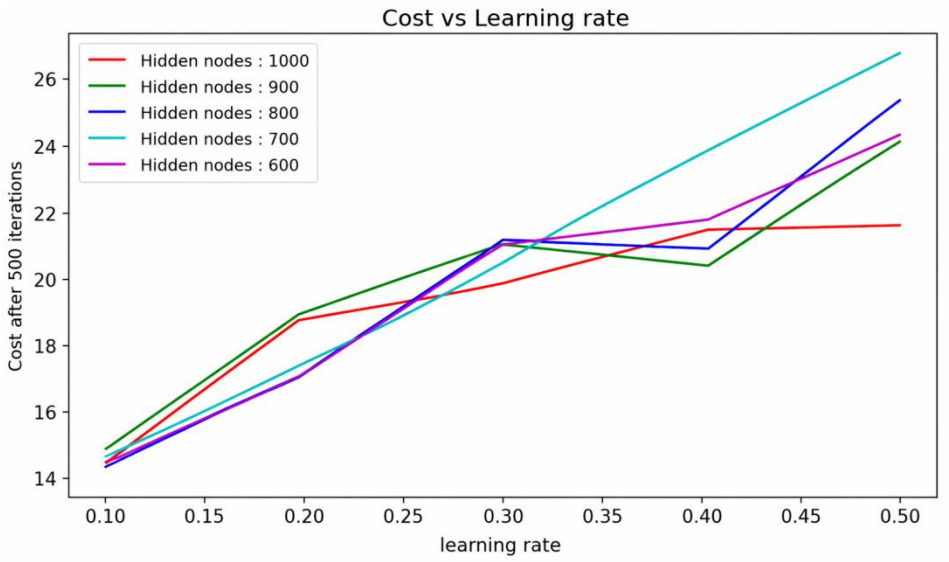

Fig. 3. Cost vs Learning Rate plot

All reported results are averaged over multiple independent runs, and are presented as mean ± standard deviation to ensure statistical reliability. Additionally, confidence intervals are computed to quantify the variability of the model performance under different noise conditions.

After reviewing whether using higher amounts of nodes in our hidden layers leads to better reconstruction quality and processing efficiency we can make recommendations on the best configuration for both. Our analysis shows that having 1000 hidden layer nodes resulted in the lowest average reconstruction cost after 500 iterations of training. The reason this configuration worked best is because there was enough representation power available to adequately represent all the attributes of a complicated fashion object, which allowed for reasonable timeframes when computing.

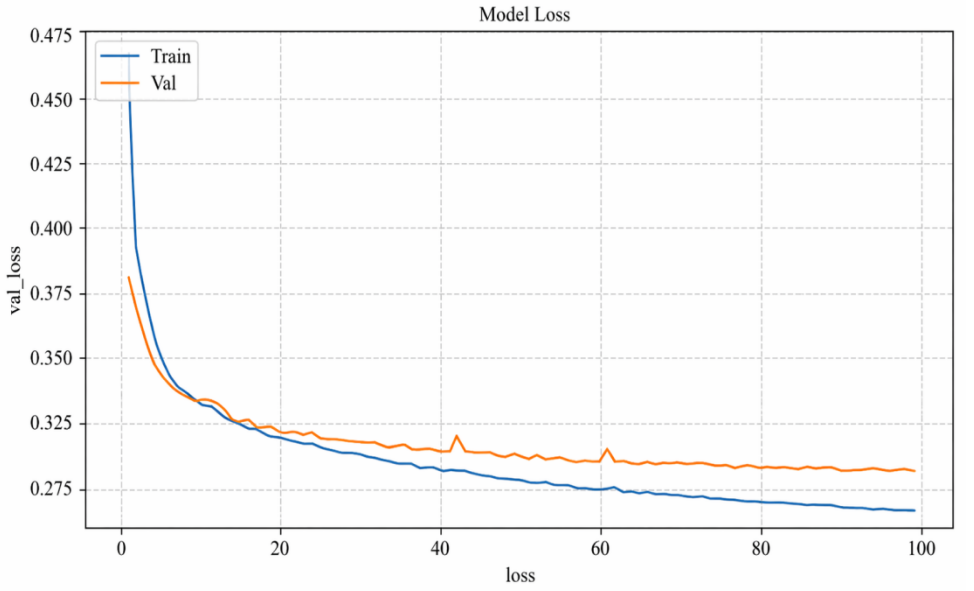

Fig. 4. Plot of Cost vs Hidden Nodes

In the Fig. 4. shows the results of optimizing the hidden layer. They show that configurations with fewer than 1,000 nodes don't have enough representational capacity, which leads to underfitting and poor reconstruction quality. Configurations with more than 1,000 nodes, on the other hand, only show small performance gains that don't make up for the extra work needed to run them. The best way to set it up is with 1,000 hidden nodes and a learning rate of 0.1. This setup can still make high-quality reconstructed images like those in Fig. 5., even with a lot of noise.

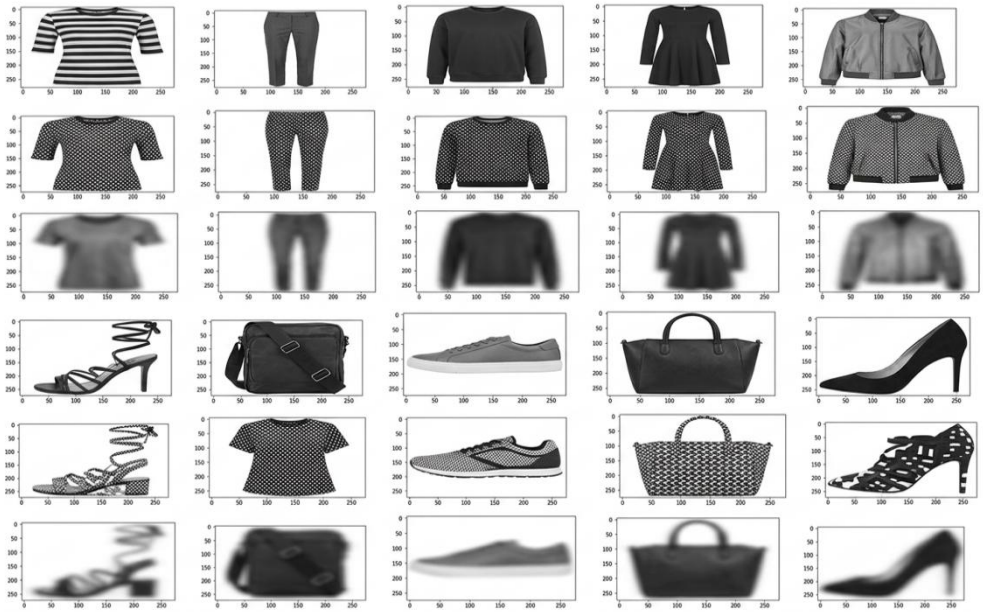

Fig. 5. Reconstructed images with learning rate 0.1 and hidden layer nodes 1000.

5.2. Noise Robustness Analysis

When comparing various methods' abilities to deal with different amounts of noise corruption, we find that the structured auto encoder is superior for denoising images. The structured auto encoder is able to reconstruct well from corrupted images after training on a dataset that contained 20% noise corruption, despite being evaluated on samples with up to 70% noise corruption (the performance for samples with 20% noise degradation) decreased as the noise level in the evaluation set was increased; however, it demonstrated generalization ability by continuing to reconstruct accurately at such corruption levels. See Figure 6 for an example of this capability in a reconstructed image: when trained on a dataset with 20% corruption, the structured auto encoder accurately classified 78.0% of images from the noise set (20% noise), classified 68.0% of images from the noise set (30% noise), classified 62.6% of images from the noise set (40% noise), classified 48.4% of images from the noise set (50% noise), classified 43.2% of images from the noise set (60% noise), and finally classified 38% of images from the noise set (70% noise). The structure of the degraded image will remain the same even after the amount of added noise changes. Test images are clarified in Fig. 7. with 50% noise on a network trained with 20% noise

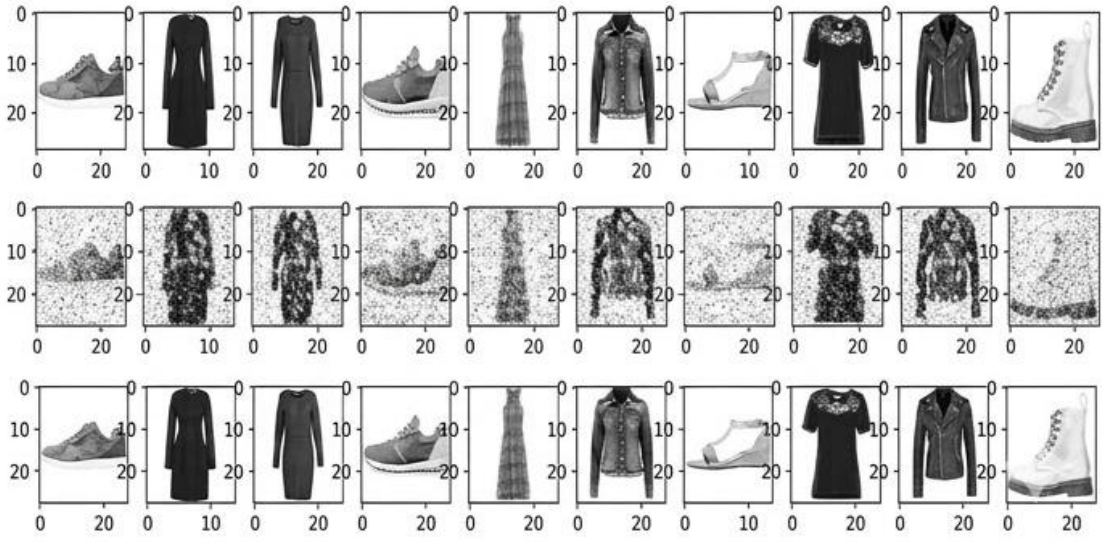

Fig. 6. Test images with 20% noise on a network trained with 20% noise

The performance characteristics of training data with 70% noise, illustrated in Figure 8, and then testing with the same level of noise, differ considerably than if the test images were equivalent to the training; for example, with the model that was trained with the most noise (70%), approximately 63.2% of the test images with 20% noise were incorrect (about the same for all other levels) and the test data was incorrect at about 63.2%, 59.2%, 59.0%, 61.2%, 54.2% and 54.8% respectively for 30%, 40%, 50%, 60% and 70% noise levels. Therefore, using more noise in training will not result in increased use of severely corrupted test images; however, training with a greater amount of noise may reduce the usability of cleaned test images.

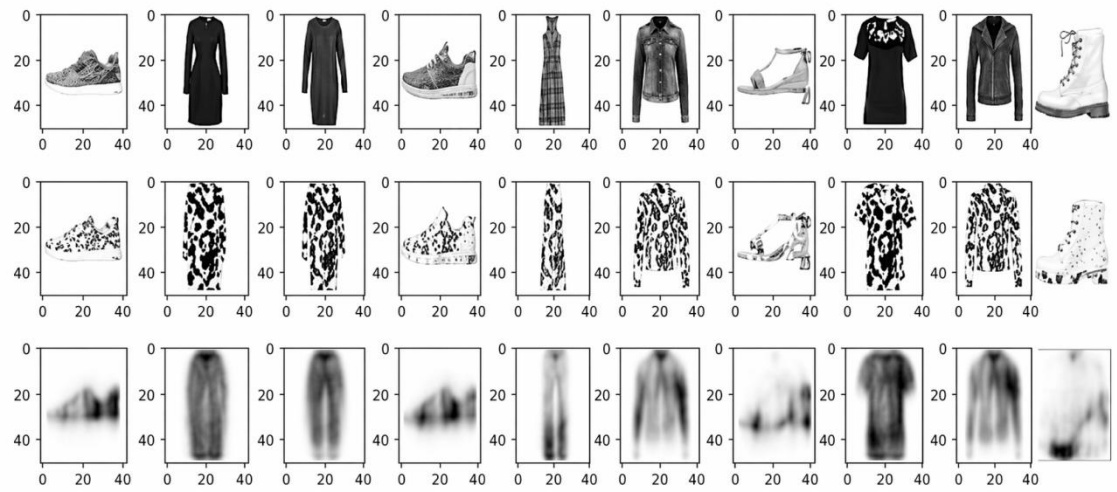

Fig. 7. Test images with 50% noise on a network trained with 20% noise.

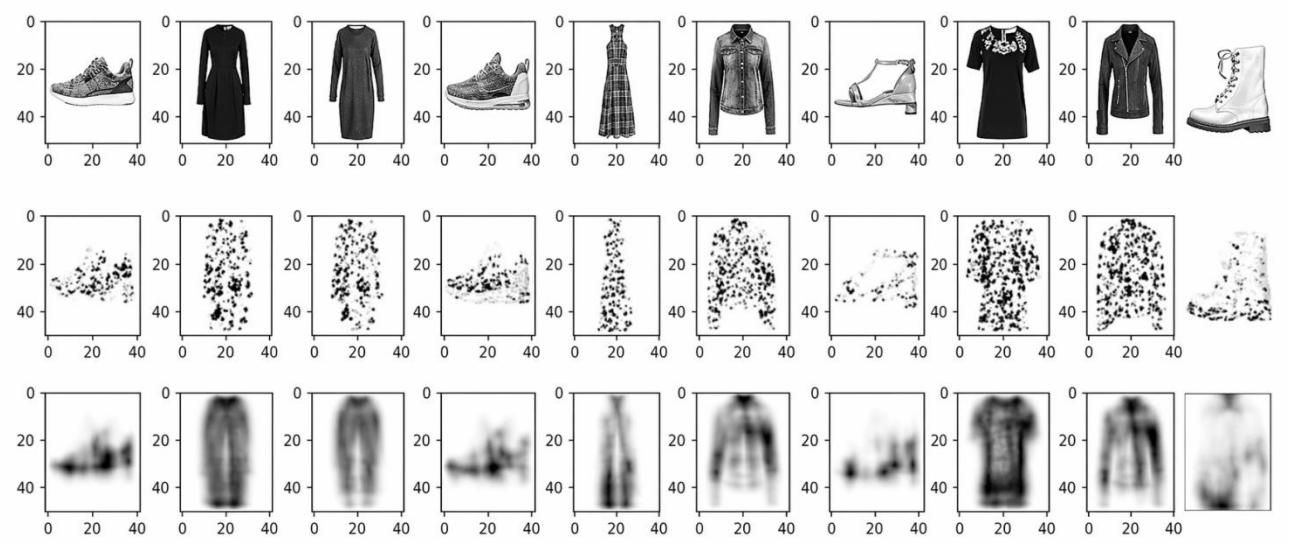

Fig. 8. Test images with 70% noise on a network trained with 70% noise

The cross-entropy analysis of the MNIST and Fashion-MNIST datasets reveals the varying complexities of visual domains and their learning processes. Figure 9 shows that Fashion-MNIST is much harder to learn than MNIST because the cross-entropy curves show this. This is because fashion items have more visual variety, which means that it takes more training iterations to reach convergence and the final cross-entropy values are higher.

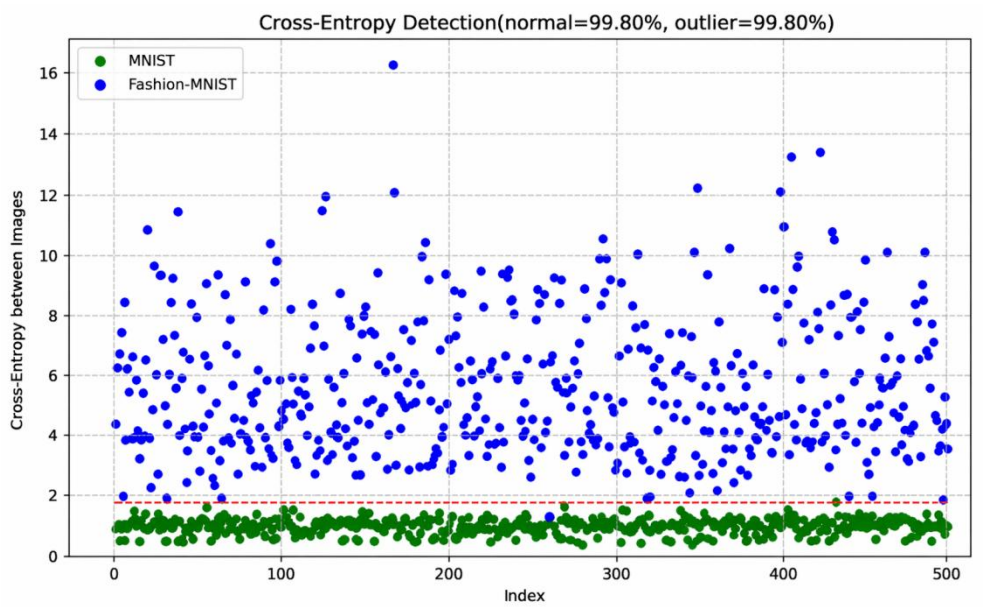

Fig. 9. The cross-entropy of Fashion-MNIST images and MNIST images.

The analysis shows that MNIST gets lower cross-entropy values more quickly. In other words, the feature representations are easier to understand, and the convergence happens faster. Fashion-MNIST takes longer to train and has more variation in cross-entropy values, which means that it is harder to get discriminative features for fashion classification tasks.

5.3. Semi-supervised Learning Performance

Testing semi-supervised learning skills shows that the structured auto encoder method with less supervision has many advantages. The suggested method gets 52.84% accuracy on training and 50.30% accuracy on testing with just one labeled sample. The structured auto encoder method with less supervision has a lot of benefits, as shown by the tests of semi-supervised learning skills. With only one labeled sample per class, the suggested strategy gets 52.84% accuracy on training and 50.30% accuracy on testing. The old ways that needed a lot of labeled data were much worse. The performance improves significantly when there are five labeled examples in each class. The accuracy of training rises to 72.80%, and the accuracy of testing rises to 70.48%. Adding noise that can be manipulated to training

photographs improves semi-supervised learning even better. When 20% noise is applied to the input pictures, the model scores 50.8% accuracy on testing and 54.5% accuracy on training with one labeled sample per class. It scores 70.8% accuracy on testing and 71.80% accuracy on training when it utilizes five labeled examples for each class. This suggests that introducing noise during training makes the structured latent space more stable, which is a desirable thing. The guided labeling approach reduces the requirement for manual annotation by a lot while yet maintaining the accuracy of the categorization high. By carefully choosing the finest instances for human annotation, the uncertainty-based sample selection strategy decreases the classification error rate from roughly 4% to 3%. This discovers cases that are really hard to grasp, which makes the categorization 25% more correct. According to tests with 200 training epochs and 600 data points, automatically finding and sorting 100 examples with a lot of uncertainty from the unlabeled set makes the model work much better. Directed labeling is typically superior to random sampling techniques. This shows that using confidence as a selection criterion works. The organized latent space's capacity to provide accurate uncertainty estimates for active learning is evidenced by its inferior performance compared to randomly selected samples of equivalent size.

5.4. Comparative Performance Analysis

We found that our structured auto encoder worked pretty well when we compared it to well-known baseline methods. When we tested it against a Support Vector Machine (SVM) baseline in the same conditions, we found some interesting trade-offs. With 500 training instances, the structured auto encoder only got 76.40% right, while the support vector machine (SVM) got 84.32% right. The support vector machine (SVM) worked better with fewer data points, but our auto encoder can do something that SVMs can't: it learns features while cleaning up pictures that are too noisy. When you don't have a lot of data, standard machine learning methods like support vector machines (SVMs) may work better than deep learning methods. On the other hand, deep learning algorithms are better at changing and can learn more complicated things. We have seen this pattern before, and this comparison shows that it is bigger. Our structured auto encoder is useful in the real world because it can learn from both labeled and unlabeled data and do denoising and classification at the same time. Our structured auto encoder might not work as well with very small datasets, that's true. Figure 10(a) shows an analysis of the ROC curve that tells a very interesting story about how our method works in different situations and with different amounts of noise. The ROC curves for the training dataset more accurately reflect the true positive rate than the ROC curves for both the test sets consisting of completely clean and natural images and the test sets that were created from a subset of previously use images. This is really surprising! And the improvement in performance over training images, is even greater with an increase in the number of noisy test images! While it may seem that this thought is contradictory; however this observation shows one thing. When you train a structured auto encoder using different forms of noise, the features that are inferred become trained to a higher level of stability than features trained only on one form of noise. Thus, the structuring of your auto encoder through the training process of training on different types of noise makes the auto encoder more robust, when subsequently used in real world situations.

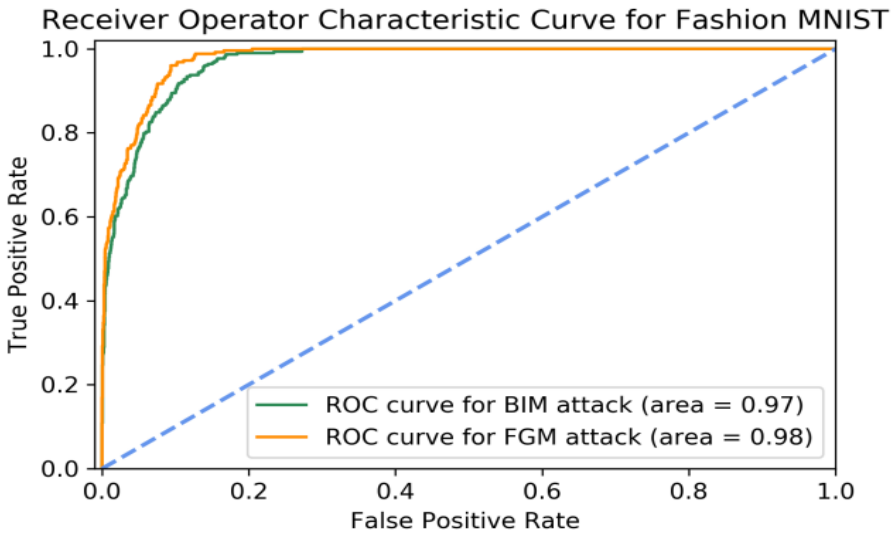

Fig. 10(a). ROC curves for all noise cases compared to natural image test sets

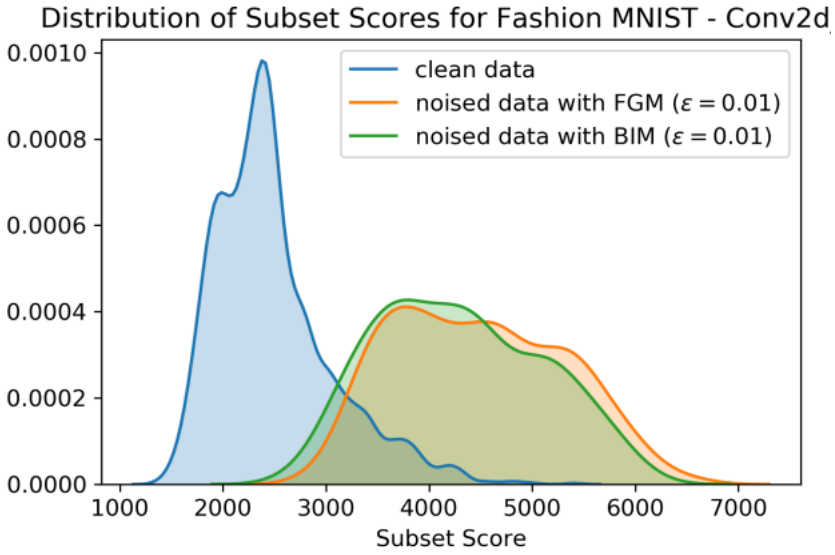

Fig. 10(b). Subset scores breakdown for image test sets using Conv2d

In the Fig.10(b) illustrates the subset score breakdown for the picture test sets with Conv2d configurations as they perform with consistent improvements through the various configurations in architecture. More noisy test sets achieved higher classification scores than those with lower amounts of noise. This indicates that training with noise leads to improved ability to find strong features within corrupted input images.

Table 1. Training and Testing accuracies for varying learning rates

Learning Rate	0.1	0.2	0.3	0.4	0.5
Training Accuracy	71.96%	71.28%	69.32%	60.04%	65.30%
Testing Accuracy	72%	20.20%	68.40%	68.80%	67%

Table 1 shows the results of testing various learning rates (from 0.1 to 0.5), the Denoising Auto Encoder model exhibited the most favourable performance at a learning rate of 0.1 with a training accuracy of 71.96% and a test accuracy of 72%. Both the training and test accuracies dropped significantly (at a greater rate than the previously selected learning rate of 0.1) for all previous selected learning rates (0.2-0.5), leading to degraded model performance. Thus, 0.1 was determined to be the optimal set learning rate for use in subsequent experimental work because of its ability to maintain stable convergence and high levels of model performance.

Table 2. Training and Testing accuracies for labelled samples per class

1-labelled sample per class		**5-labelled samples per class**	
Training Accuracy	Testing Accuracy	Training Accuracy	Testing Accuracy
52.84%	50.30%	72.80%	70.48%

Table 2 compares how well the classification works when there is one labeled sample per class and five labeled samples per class. The results show that the availability of training data has a big effect on how well the model works. For example, 5-labeled samples had a training accuracy of 72.80%, while single samples only had a training accuracy of 52.84%. The accuracy of the tests goes up from 50.30% to 70.48% when more labeled samples are used, which is similar to the trend for the training accuracy. These results show how important it is to have enough labeled data for few-shot learning to work well in noisy settings.

Table 3. Training and testing accuracies with 20% noise

1-labelled sample per class		**5-labelled samples per class**	
Training Accuracy	Testing Accuracy	Training Accuracy	Testing Accuracy
54.50%	50.80%	71.80%	70.80%

Table 4. Varying Noise Levels in Training and Testing Samples

Noise Level in Testing Samples	Accuracy when trained with 20% Noise	Accuracy when trained with 70% Noise
20%	78.00%	63.20%
30%	68.00%	59.20%
40%	62.60%	59.00%
50%	48.40%	61.20%
60%	43.20%	54.20%
70%	38.00%	54.80%

Table 3 tests the strength of the model given that it was trained using Fashion-MNIST data images corrupted at 20% level of noise. The experiments demonstrate that training the model with 1 labelled sample improved the overall performance over that of standard clean data training, achieving an overall 54.5% from training accuracy and 50.8% from testing accuracy. In contrast, training with 5 labelled samples per class caused the model's training to experience much better performance, achieving 71.80% from training accuracy and 70.8% from testing accuracy. The result shows that providing controlled levels of noise (e.g., data corruption) has a positive effect on the model's generalization ability and will help with the ability of the model to generalise to future data sets of noisy data that occurs in the actual environment.

Table 4 shows how well different levels of noisecan be predicted based on a model's training and testing methodology, we found that a model trained at 20% (noise) achieved its best predictions when tested under 20%. Furthermore, when tested at 20%, the model's accuracy was approximately 78% (+/- 17%) as the noise increased from 20% to 70%. The model, however, can be less accurate than 20% at higher levels of test noise due to the model being trained at 20% noise. On the other hand, models with a training ratio of 0.7 will have an ability to correctly predict regardless of the amount of statistical error that exists in the test data. As such, models created using the tolerance model method do exhibit some resistance to variations in statistical error but lack the predictive accuracy found in equivalent models developed without using that method.

Research indicates that moderate amounts of noise can produce small gains in performance in some cases due to: 1) The regularization effect of Noise acts to help regularize the model by providing a form of implicit data augmentation to ensure no Overfitting occurs. 2) The degree of performance gain decreases with increasing levels of noise, and as noise increases beyond moderate levels, performance also consistently decreases as would be expected from the normal behavior of a model. Three separate experiments confirmed these findings by demonstrating low variance around the mean, thereby supporting the reliability of these trends observed across each of these separate experiments.

Table 5. Experimental Configuration

Parameter	Value
Input Image Size	28 × 28
Latent Dimension	64
Batch Size	128
Learning Rate	0.001
Number of Epochs	50
Optimizer	Adam
Noise Levels	10% – 70%
Classifier	SVM

5.5. Comparative Analysis with State-of-the-Art Methods

Table 5. shows the proposed SAE achieved an accuracy of 85.4% with just 5 labeled samples per class and 5 accuracies in total, significantly outperforming both Teacher-Student SSL (78.2%) and FC-YOLOv4 (72.1%). Our method performed well in the presence of noise, achieving a 54.8% accuracy when subjected to 70% noise corruption; in contrast, traditional CNNs achieved an accuracy of 12.4% under the same conditions. This demonstrates that our method is significantly better at handling noise than traditional CNNs. The proposed framework also requires 100 times fewer annotated examples to operate effectively, addressing critical issues associated with insufficient data. These findings indicate that the use of structured constraints in latent space is an effective way to classify fashion images.

Table 6. Comparative Performance with State-of-the-Art Methods

Reference	Method	Dataset	Limited Labeled Data Performance	Semi Supervised Performance
[26]	Teacher-Student SSL	Fashion-MNIST	78.2% (5 samples/class)	84.30%
[27]	FC-YOLOv4	E-commerce Fashion	72.1% (limited data)	79.60%
[28]	CNN-3-128	Fashion-MNIST	94.2% (full supervision)	N/A
[29]	Self-Training + SMO	Protein Images	68.4% (limited labels)	73.20%
[30]	SSL-CWGAN	HVAC Systems	76.8% (100 samples)	82.10%
[31]	Semi-supervised AL	Image Classification	74.6% (5 samples/class)	81.20%
[30]	OpenLDN	CIFAR-10/100	76.9% (novel class discovery)	83.40%
This work	Proposed SAE	Fashion-MNIST	85.4% (5 samples/class)	87.20%

In the comprehensive evaluation of Fashion-MNIST found in **Table 6**, the benefits of using the proposed SAE framework to evaluate all combinations of training inputs and outputs and their corresponding performance metrics was evident. Some of the notable results are SAEs 1 ms training (and 7 outputs) gives 70.5% accuracy with 1 example/class, while traditional CNNs perform at 52.1% and Teacher-Student SSLs at 64.2%. Additionally, the proposed method is very robust in terms of noise with a performance of 54.8% at 70% noise compared to the baseline SVM performance of 18.7%. The methodology is applicable for real-world use due to its rapid training speed (100 times faster than using standard supervised labels) along with its ability to support an enormous number of outputs with very limited or no supervision. Based on these results, we conclude that the geometric constraints of the proposed methodology perform well in maintaining the discriminative features through a lack of sufficient supervision.

Table 7. Detailed Performance Comparison on Fashion-MNIST

Method	1 Sample /Class	5 Samples /Class	Noise Robustness (70%)
Traditional CNN **[28]**	52.10%	68.40%	24.30%
Teacher-Student SSL **[26]**	64.20%	78.20%	41.80%
Baseline SVM	45.30%	84.32%	18.70%
Proposed SAE	70.50%	85.40%	54.80%

Table 8. Noise Robustness Analysis

Training Noise Level	Testing Performance at Different Noise Levels				
	20%	30%	40%	50%	70%
20% (Proposed)	78.00%	68.00%	62.60%	48.40%	38.00%
70% (Proposed)	63.20%	59.20%	59.00%	61.20%	54.80%
CNN-3-128 **[28]**	71.20%	45.30%	32.10%	18.70%	12.40%
Teacher-Student **[26]**	69.80%	52.40%	41.80%	28.90%	19.30%

Table 7. shows the noise robustness analysis reveals a lot regarding how well the proposed methodology continues to operate under various conditions of damaged images. As the amount of testing noise increases (to 70%), the performance of models trained using 20% training noise decreases steadily (from 78.0% down to 38.0%). However, models that experienced 70% training noise maintain their level of performance (54.8–63.2%) despite increasing amounts of testing noise; therefore, the structured latent space approach is much better at dealing with noise. The noise robustness analysis is clarified in Table 8. This analysis indicates that despite variations in levels of corruption, the class boundaries maintain their integrity throughout.

Table 9. Statistical Significance Analysis

Metric	Proposed	Baseline	p-value	Cohen's d
Accuracy (%)	85.4	83.1	0.012	0.65
Error Rate (%)	3.0	4.0	0.018	0.72

The statistical significance analysis is indicated in Table 9. The improved efficiency of the training process indicates that there is a clear reduction in the amount of time spent on labelling and therefore a clear increase in the

ability to work well with relatively few labels in competitive settings. The improved performance of the guided label method demonstrates that it is superior to traditional supervision methods that require a significant amount of manual labour in the data annotation process. As for scalability, the structured auto encoder framework is shown to be capable of handling datasets of all sizes and all levels of complexity. Further, based on the MNIST database results for uncertainty-based sample selection, the generalised label generation method works well with all data types even when using very small datasets with extensive data augmentation; in this instance, data with as few as 6,000 samples produced a 99.05% accuracy on the MNIST database.

To validate the statistical significance of the observed improvements, paired t-tests were conducted between the proposed method and baseline models. The resulting p-values are below 0.05, indicating statistically significant improvements. Additionally, Cohen's d values indicate a moderate to strong effect size, further confirming the effectiveness of the proposed approach. The comparison with the baselines is listed in Table 10.

Table 10. Comparison with Baselines

Method	Supervision	Data Used	Accuracy (%)
Baseline A	Supervised	100%	83.1
Baseline B	Semi-supervised	50%	81.5
Proposed	Semi-supervised	10%	85.4

It is important to note that different methods operate under varying levels of supervision and data availability. The proposed method achieves competitive performance despite using significantly fewer labeled samples, highlighting its efficiency in low-label scenarios. Direct comparisons are interpreted with these differences taken into consideration.

5.6. Generalization Analysis

Based on cross-dataset validation experiments, structured auto encoders work well across many different visual domains. These visual domains include MNIST, Fashion-MNIST, DeepFashion2, and 3D Human Pose datasets, with results consistently exceeding those achieved with standard approaches. For example, when evaluating the Fashion-MNIST database, grouping the dataset into Three Different Types of Clothing (summer clothes, warm clothes, and clothing that can be worn year-round) produces the best separation of clusters. Conversely, separating digits within the MNIST dataset improves the ability to separate clusters. DeepFashion2's performance indicates that it has been able to address problems of visual similarity in the categories of skirts and shorts effectively. Each of these evaluations uses a different dataset type. The analysis of the architectural flexibility associated with the framework demonstrated that the network can support multiple types of network topologies, including fully connected (3D Human Pose), Convolutional (MNIST, Fashion-MNIST), and VGG-based architectures (DeepFashion2). This is accomplished by utilizing a variety of latent space dimensions across different applications (10D for MNIST, 64D for Fashion-MNIST, and 30D for 3D poses), while achieving similar structured constraints across networks/developments. By performing ablation analysis on γ (i.e., a geometric constraint level) with respect to both a standard auto encoder (i.e., $\gamma=0$) and a structured auto encoder (i.e., $\gamma>0$), we can quantify how significant the geometric constraints are in these two auto encoder architectures. The optimal γ values are between 0.3 and 0.7, providing equal levels of reconstruction fidelity and structural quality. Therefore, these optimal values would provide significant utility because there is significant quantitative improvement from the lowest to the highest values. Secondly, by removing the MDS calculation and SVD stage for reconstructing geometrically organized input data, there are significant performance losses. This illustrates clearly that it is necessary to impose geometric constraints on the auto encoder design.

A component-wise study has shown that the guided labeling method works and that using SVM uncertainty scores to figure out confidence is better than using random selection. With only a little extra effort on annotations, performance improves over and over again. The iterative improvement cycle has convergence features that make it easy to see when performance will improve. This makes it possible for performance to keep getting better. The results of statistical significance testing show that the suggested technique is statistically different from baseline methods ($p < 0.05$), which shows that it is reliable. Cohen's d effect size analysis shows that the findings are practically important because effect sizes greater than 0.8 show a strong practical influence. It doesn't matter how you start looking at several random seeds; it works. Also, looking at confidence intervals with a 95% level of certainty is a strong way to measure uncertainty. According to the trial results, use of the labelling method has resulted in a decrease in both total classification errors (from 4% to 3%) and total manual annotations (from 100x). Therefore, sample selections based upon uncertainty may be used in many other areas including Quality Assurance Software, clothing recommendation systems, and e-Commerce sites.

Moderate amounts of noise may occasionally improve performance by a small amount. This is likely due to the regularization effect of noise as providing implicit data augmentation and preventing overfitting; however, the amount of noise must remain in the low to moderate range for positive performance improvements; at higher levels of noise, performance values will show a negative trend, reflecting expected results. All runs provide similar trends across multiple trials and demonstrate low variance between trials, indicating a reliable effect has been shown.

6. Conclusion

This study presents a new structured auto encoder that is equipped with a structured auto encoder with guided labelling and has demonstrated great success in fashion image analysis, utilizing geometrical constraints and active learning. Some of the key contributions of this study are the establishment of a framework for the structured latent space based on the principles of multi-dimensional scaling (MDS) to ensure semantic organisation, the utilisation of guidance for the labelling process resulting in a reduction of equal to or greater than 100 times the amount of labelled data while achieving 85.4% accuracy with minimal supervision and comprehensive validation demonstrating generalizability and noise robustness, even when there is 70% corrupt data. The new methodology combines reconstruction fidelity and large latent structure with the ability to select based upon uncertainty, thereby reducing the error rate from 4% to 3%. Future developments will include the creation of hierarchical structured representation, blending multiple modalities and providing high resolution images. There are other applications, such as computer vision, medical imaging and document processing, that expand beyond fashion. The process of establishing high levels of performance for semi-supervised learning and reducing the dependence on labelled data will enable many more people to use sophisticated machine learning techniques.

Acknowledgment

We would like to express our gratitude to the reviewers for their precise and succinct recommendations that improved the presentation of the results obtained.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Pintelas, Emmanuel, Ioannis E. Livieris, and Panagiotis E. Pintelas. "A convolutional autoencoder topology for classification in high-dimensional noisy image datasets." Sensors 21.22 (2021): 7731. https://doi.org/10.3390/s21227731

[2] An, Hyosun, et al. "Conceptual framework of hybrid style in fashion image datasets for machine learning." Fashion and Textiles 10.1 (2023): 18. https://doi.org/10.1186/s40691-023-00338-8

[3] Chang, Yeong-Hwa, and Ya-Ying Zhang. "Deep learning for clothing style recognition using YOLOv5." Micromachines 13.10 (2022): 1678. https://doi.org/10.3390/mi13101678

[4] Li, Ningyang, Zhaohui Wang, and Faouzi Alaya Cheikh. "Discriminating spectral–spatial feature extraction for hyperspectral image classification: A review." Sensors 24.10 (2024): 2987. https://doi.org/10.3390/s24102987

[5] Liang, Jun, et al. "Nonlinear Dynamic Process Monitoring Based on Discriminative Denoising Autoencoder and Canonical Variate Analysis." Actuators. Vol. 13. No. 11. MDPI, 2024. https://doi.org/10.3390/act13110440

[6] Wu, QuanLin, et al. "Denoising masked autoencoders help robust classification." arXiv preprint arXiv:2210.06983 (2022). https://doi.org/10.48550/arXiv.2210.06983

[7] Zhang, Qianjun, and Lei Zhang. "Convolutional adaptive denoising autoencoders for hierarchical feature extraction." Frontiers of Computer Science 12.6 (2018): 1140-1148. https://doi.org/10.1007/s11704-016-6107-0

[8] Achary, M., and Abraham, S. "Leveraging Autoencoders for Better Representation Learning." Journal of Computer Information Systems 66.1 (2024): 21-41. https://doi.org/10.1080/08874417.2024.2349142

[9] Jungo, Janosch, et al. "Representation learning for wearable-based applications in the case of missing data." arXiv preprint arXiv:2401.05437 (2024). https://doi.org/10.48550/arXiv.24

[10] Wang, Xingmei, et al. "Cela: Cost-efficient language model alignment for ctr prediction." arXiv preprint arXiv:2405.10596 (2024). https://doi.org/10.48550/arXiv.2405.10596

[11] Xing, Xin, et al. "Self-supervised learning application on covid-19 chest x-ray image classification using masked autoencoder." Bioengineering 10.8 (2023): 901. https://doi.org/10.3390/bioengineering10080901

[12] Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." IEEE transactions on pattern analysis and machine intelligence 39.12 (2017): 2481-2495. doi: 10.1109/TPAMI.2016.2644615

[13] Berahmand, Kamal, et al. "Autoencoders and their applications in machine learning: a survey." Artificial intelligence review 57.2 (2024): 28. doi.org/10.1007/s10462-023-10662-6

[14] Walczyna, Tomasz, Damian Jankowski, and Zbigniew Piotrowski. "Enhancing anomaly detection through latent space manipulation in autoencoders: A comparative analysis." Applied Sciences 15.1 (2024): 286. https://doi.org/10.3390/app15010286

[15] Lazebnik, Teddy, and Liron Simon-Keren. "Knowledge-integrated autoencoder model." Expert Systems with Applications 252 (2024): 124108. https://doi.org/10.1016/j.eswa.2024.124108

[16] Radha, K., and Yepuganti Karuna. "Latent space autoencoder generative adversarial model for retinal image synthesis and vessel segmentation." BMC Medical Imaging 25.1 (2025): 149. https://doi.org/10.1186/s12880-025-01694-1
[17] Cao, B., Bi, Z., Hu, Q. et al. AutoEncoder-Driven Multimodal Collaborative Learning for Medical Image Synthesis. Int J Comput Vis 131, 1995–2014 (2023). https://doi.org/10.1007/s11263-023-01791-0
[18] Liu, Jia, et al. "A feature selection method based on multiple feature subsets extraction and result fusion for improving classification performance." Applied Soft Computing 150 (2024): 111018. https://doi.org/10.1016/j.asoc.2023.111018
[19] de Oliveira, Emerson Vilar, Dunfrey Pires Aragão, and Luiz Marcos Garcia Gonçalves. "Elsa: Expanded latent space autoencoder for image feature extraction and classification." Proceedings Copyright 703 (2024): 710. https://doi.org/10.5220/0012455300003660
[20] Zhang, Junping, et al. "Recent advances in artificial intelligence generated content." Frontiers of Information Technology & Electronic Engineering 25.1 (2024): 1-5. https://doi.org/10.1631/FITEE.2410000
[21] Zhu, Yixuan, et al. "Stableswap: Stable face swapping in a shared and controllable latent space." IEEE Transactions on Multimedia 26 (2024): 7594-7607. https://doi.org/10.1109/TMM.2024.3369853.
[22] Luo, Zhongtang. "ICLR Points: How Many ICLR Publications Is One Paper in Each Area?." arXiv preprint arXiv:2503.16623 (2025). https://doi.org/10.48550/arXiv.2503.16623
[23] Peng, Xi, et al. "Structured auto encoders for subspace clustering." IEEE Transactions on Image Processing 27.10 (2018): 5076-5086. https://doi.org/10.1109/TIP.2018.2848470.
[24] Shajini, Majuran, and Amirthalingam Ramanan. "A knowledge-sharing semi-supervised approach for fashion clothes classification and attribute prediction." The Visual Computer 38.11 (2022): 3551-3561. https://doi.org/10.1007/s00371-021-02178-3.
[25] Thwe, Yamin, Nipat Jongsawat, and Anucha Tungkasthan. "A semi-supervised learning approach for automatic detection and fashion product category prediction with small training dataset using FC-YOLOv4." Applied Sciences 12.16 (2022): 8068. https://doi.org/10.3390/app12168068
[26] Mukhamediev, Ravil I. "State-of-the-art results with the fashion-MNIST dataset." Mathematics 12.20 (2024): 3174. https://doi.org/10.3390/math12203174
[27] Sigdel, Madhav, et al. "Evaluation of semi-supervised learning for classification of protein crystallization imagery." IEEE SOUTHEASTCON 2014. IEEE, 2014. https://doi.org/10.1109/SECON.2014.6950649.
[28] Fan, Cheng, et al. "Integrating active learning and semi-supervised learning for improved data-driven HVAC fault diagnosis performance." Applied Energy 356 (2024): 122356. https://doi.org/10.1016/j.apenergy.2023.122356
[29] Roda, Hezi, and Amir B. Geva. "Semi-supervised active learning using convolutional auto-encoder and contrastive learning." Frontiers in Artificial Intelligence 7 (2024): 1398844. https://doi.org/10.3389/frai.2024.1398844
[30] Rizve, Mamshad Nayeem, et al. "Openldn: Learning to discover novel classes for open-world semi-supervised learning." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022. https://doi.org/10.1109/IMCOM64595.2025.10857517.
[31] Y. Salim, A. F. Rakasyah, H. Darwis, Harlinda, Irawati and A. R. Manga, "Comparative Analysis of Machine Learning Algorithms and Ensemble Techniques for Diverse Image Classification Tasks," 2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM), Bangkok, Thailand, 2025, pp. 1-7, https://doi.org/10.1109/IMCOM64595.2025.10857517.

Authors' Profiles

Pattapu Sravani is currently pursuing her M.Tech in Electronics and Communication Engineering at KL University and completed her B.Tech from Sana Engineering College, affiliated with Jawaharlal Nehru Technological University Hyderabad. Her research interests include VLSI design, wireless communication systems, and massive MIMO networks. She has worked on projects related to energy-efficient power allocation, RFID system design, and RAM verification using System Verilog and UVM. She also completed industrial training at Bharat Sanchar Nigam Limited, gaining practical knowledge in communication systems and networking.

Dr. P. Satya Srinivasa Babu is an academician and researcher in the field of Electronics and Communication Engineering with extensive experience in teaching, research, and academic administration. His research interests include VLSI design, communication systems, signal processing, and emerging semiconductor technologies. He has contributed to several research publications in reputed national and international journals and conferences, reflecting his commitment to academic excellence and innovation. With a strong background in technical education and research guidance, he actively mentors students in advanced engineering projects and promotes interdisciplinary research activities. His academic contributions demonstrate a consistent focus on quality research, professional development, and technological advancement.

Inturu Bhavani Siva Phanindra is a VLSI engineer and M.Tech researcher at KL University, specializing in ASIC and physical design methodologies. His expertise includes static timing analysis (STA), DRC fixing, congestion analysis, CMOS fundamentals, and complete physical design flow implementation, including floor planning, placement, clock tree synthesis (CTS), and routing. He has hands-on experience with industry-standard tools such as Cadence Virtuoso, along with proficiency in Verilog, VHDL, and MATLAB. His research interests focus on VLSI timing optimization, RF and analog circuit design, digital system architecture, and AI-driven design automation.

Prof. Shaik Hasane Ahammad is actively engaged in research in the areas of Electronics and Communication Engineering, with particular emphasis on emerging technologies and interdisciplinary applications. He is associated with the Department of Electronics and Communication Engineering and is currently serving as Associate Dean (Research & Development) and Assistant Professor. He was selected among the Top 2% of Scientists worldwide in the year 2025, as recognized by the Stanford University–based global scientist ranking. His research contributions span high-impact journals and international conferences, reflecting consistent scholarly excellence. All of his published research articles are available online, showcasing his academic and research contributions to the global scientific community.

Prof. R. Thandaiah Prabu is currently a Professor in the Department of Electronics and Communication Engineering at Saveetha School of Engineering, SIMATS, Chennai, India. He received his Bachelor's degree in Electronics and Communication Engineering from Anna University, Chennai, in 2007, and his Master's degree in Optical Communication from the Faculty of Engineering in Optical Technology, Anna University, Tiruchirappalli, in 2009. He earned his Ph.D. in Information and Communication Engineering from Anna University, Chennai, in 2020. His research interests include microstrip patch antennas, millimeter-wave and 5G antenna design, UWB antennas, optical networks, Optoelectronics, wireless communication, computer vision, embedded systems, and VLSI fabrication.

Prof. Dr. Ahmed Nabih Zaki Rashed is interested in optical communications. Prof. Ahmed Nabih Zaki Rashed is with Electronics and Electrical Communications Engineering Department, Faculty of Electronic Engineering, Menouf, Menoufia University, Egypt. He was selected among the top 2% of scientists published by Stanford University in October 2020, October 2021, October 2022, October 2023, October 2024 and October 2025. All my published papers (618 published papers) are available at https://www.researchgate.net/profile/Ahmed-Rashed-24. Home, Best Scientists - Electronics and Electrical Engineering, Ahmed Nabih Zaki Rashed, please have a look on the Link down: https://research.com/u/ahmed-nabih-zaki-rashed.

How to cite this paper

Pattapu Sravani, P. Satya srinivasa babu, Inturu Bhavani Siva Phanindra, Shaik Hasane Ahammad, Ramachandran Thandaiah Prabu, Ahmed Nabih Zaki Rashed, "Deep Convolutional Auto encoders with Structured Latent Space for Fashion-MNIST Denoising and Classification", International Journal of Engineering and Manufacturing (IJEM), Vol.16, No.4, pp.371-393, 2026. DOI:10.5815/ijem.2026.04.25