Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 14
TPCEE 2026

FuzzyVAE: Quality-Aware Fuzzy Fusion for Fashion Image Classification

Jiahao Wen

College of Computer Science Sichuan University Chengdu, China

2133739482@qq.com

Abstract. When the test images are affected by noise, occlusion or other quality variations, robust fashion image classification remains challenging. This paper proposes a quality-aware fuzzy fusion framework called FuzzyVAE for classification on the Fashion-MNIST dataset. This method combines classifiers based on original intensity and HOG, and uses the variational autoencoder (VAE) for input quality estimation. The reconstruction error of VAE and image roughness are jointly used to measure the degree of sample degradation, and the fuzzy membership function dynamically adjusts the contribution of the two feature views. Unlike static fusion, the proposed method adjusts the fusion weights according to the specific situation of each input image. Experimental results show that FuzzyVAE achieves an accuracy of 86.0% and a macro F1 score of 85.56% on clean images. More importantly, in the presence of Gaussian noise, random masking and blurring, it achieves an average damaged accuracy of 78.71% and a maximum comprehensive score of 82.36%. The fuzzy gate analysis further confirms that as the degree of image degradation increases, the model gradually reduces the contribution of unreliable structural features. These results indicate that FuzzyVAE provides a simple and effective solution for balancing classification performance and robustness in the presence of image condition variations.

Keywords: Fashion-MNIST; variational autoencoder; fuzzy fusion; image classification.

1. Introduction

Fashion image classification is a fundamental task in computer vision with potential applications in product recognition, smart retail, visual search, and automated inventory management. Compared to handwritten digit recognition, fashion images typically contain more complex shape structures and exhibit stronger visual similarities across categories. Fashion-MNIST was introduced as a more challenging alternative to the original MNIST benchmark dataset, comprising 70,000 (28×28) grayscale images from ten fashion categories [1]. Although modern machine learning models achieve high classification accuracy on clean images, their performance significantly degrades when test samples are corrupted by common disturbances such as noise, blur, or local information loss [2]. Therefore, maintaining reliable classification under varying image quality remains an important challenge.

Variational Autoencoders (VAEs) provide a probabilistic framework for learning compact representations of high-dimensional observations [3]. By combining reconstruction optimization with latent space regularization, VAEs simultaneously learn an encoder for latent representation and a decoder for image reconstruction. The β-VAE further introduces a weighting factor to balance reconstruction fidelity and latent regularization [4]. Most VAE-based research has focused on image generation, representation learning, or latent space modeling. However, generative models can also provide information about whether input data aligns with the learned data distribution. As a result, VAE-based distribution assessment has been applied to detect anomalies or out-of-distribution inputs [5]. This property offers a useful foundation for pre-classification sample quality evaluation.

Different feature representations may respond differently to image degradation. Raw grayscale pixels retain direct intensity information, while Histograms of Oriented Gradients (HOG) describe local gradient directions and structural features [6]. These representations provide complementary information for fashion image recognition. However, their reliability does not remain consistent

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

across all samples. Structural representations that perform well on clean images may lose effectiveness when local gradients are heavily distorted by noise or masking. Traditional score-level fusion typically combines classifiers using fixed weights, which cannot explicitly adapt to such sample-specific variations. Fuzzy ensemble learning offers an alternative mechanism by integrating multiple classifiers through flexible membership-based decision rules [7], making it suitable for modeling gradual changes in feature reliability.

Based on these observations, this paper proposes a quality-aware fuzzy fusion framework called FuzzyVAE for robust fashion image classification. The method combines classification based on raw intensity and HOG features with a VAE-guided quality assessment mechanism. Reconstruction information obtained from the VAE is combined with image roughness to estimate the degradation state of each input. A fuzzy membership function then converts this quality information into sample-specific fusion weights, allowing the contribution of different feature representations to vary according to observed image conditions. Unlike traditional static fusion methods, FuzzyVAE achieves adaptive decision fusion without introducing complex additional gating networks.

The main contributions of this work are threefold. First, VAE reconstruction information is utilized as a sample quality indicator beyond its conventional roles in image generation or dimensionality reduction. Second, we propose a lightweight fuzzy quality-aware mechanism for dynamically fusing complementary intensity and structural representations. Third, we conduct systematic evaluations of the proposed method on both clean and degraded Fashion-MNIST datasets including Gaussian noise, random masking, and image blur to assess its classification accuracy and robustness to degradation.

2. Methodology

2.1. Dataset and Preprocessing

Fashion-MNIST is used to evaluate the proposed method [8]. The dataset contains ten clothing categories, including T-shirt/top, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, and ankle boot. Each sample is represented as a 28×28 grayscale image. In the current experimental setting, 1,000 labeled training samples and 1,000 independent test samples are used. All pixel values are normalized from $[0, 255]$ to $[0, 1]$.

The training set is further divided into an optimization subset and a validation subset using a stratified 80/20 split. Therefore, 800 samples are used for model optimization and 200 samples are retained for validation. Stratified sampling is adopted to preserve the class distribution, and the random seed is fixed to 42 for reproducibility. Fig. 1 provides the class distribution of the collected dataset.

Two complementary feature views are extracted from each image. The first is the raw-intensity view, in which the normalized image is flattened into a 784-dimensional vector. The second is the structural view represented by histogram of oriented gradients (HOG). For HOG extraction, the number of orientations is set to 9, the cell size is 4×4 pixels, the block size is 2×2 cells, and L2-Hys normalization is applied. The raw-pixel representation mainly preserves global intensity and contour information, while HOG focuses more on local edge and structural characteristics.

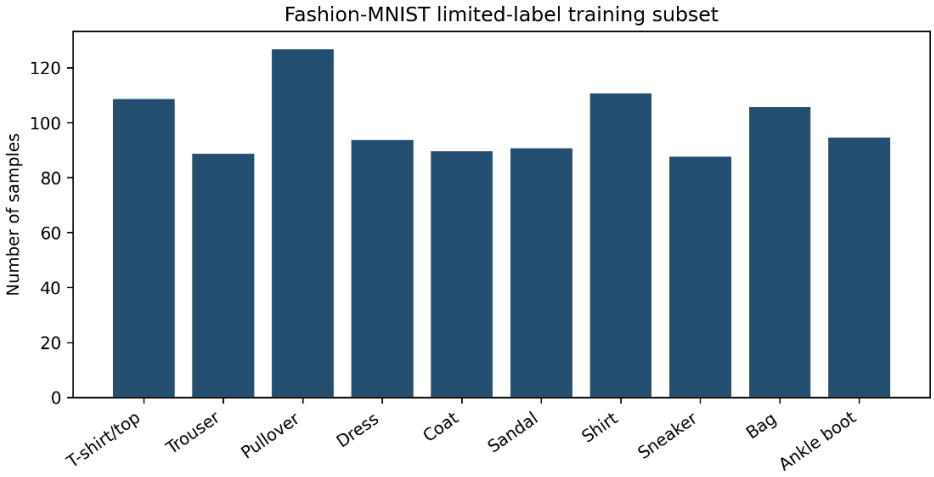

Figure 1. The class distribution of the collected dataset.

2.2. Multi-View Classification Models

Two RBF-kernel support vector machines are used as the main classification branches [9, 10]. For the raw-intensity branch, the penalty parameter is set to $C = 10$, and the kernel coefficient γ follows the scale strategy. For the HOG branch, the extracted HOG features are first standardized to zero mean and unit variance, followed by an RBF-SVM with $C = 3$ and $\gamma = \text{scale}$. Probability estimation is enabled for both classifiers to support decision-level fusion.

For an input image x, the class-probability vectors produced by the raw-pixel and HOG branches are respectively written as

$$p_r(x) = \left[p_{r,1}, (x), p_{r,2}, (x) \ldots p_{r,C}(x)\right] \tag{1}$$

and

$$p_h(x) = \left[p_{h,1}, (x), p_{h,2}, (x) \ldots p_{h,C}(x)\right] \tag{2}$$

where $C = 10$ is the number of classes.

Several additional models are included as baselines. The Extra Trees classifier contains 400 decision trees, adopts square-root feature sampling, and uses balanced class weighting. The multilayer perceptron contains two hidden layers with 128 and 64 neurons. Its input is standardized before training, with an L_2 regularization coefficient of 2×10^{-4}, a maximum of 250 iterations, early stopping, and a validation fraction of 0.15. A VAE-Latent SVM is also constructed by using the latent mean generated by the trained VAE as the input feature. These latent features are standardized and then classified using an RBF-SVM with $C = 30$.

2.3. Variational Reconstruction-Based Quality Estimation

A lightweight β-VAE is introduced as a generative quality estimator rather than as the final classifier. The encoder consists of two fully connected layers, $784 \to 256 \to 128$ followed by two independent linear layers that estimate the latent mean μ and log-variance σ^2. The latent dimension is set to 16. The latent representation is obtained using the reparameterization trick:

$$z = \mu + \sigma \odot \epsilon\,,\ \epsilon \sim \mathcal{N}(0, I) \tag{3}$$

The decoder has a symmetric structure,$16 \to 128 \to 256 \to 784,$ with ReLU activation functions in the hidden layers and a sigmoid activation in the output layer.

The VAE objective combines reconstruction loss and KL divergence regularization:

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{rec}} + \beta \mathcal{L}_{\text{KL}} \tag{4}$$

where

$$\mathcal{L}_{\text{KL}} = -\frac{1}{2}\sum_{j=1}^{d_z}\left(1 + \log \sigma_j^2 - \mu_j^2 - \sigma_j^2\right) \tag{5}$$

Based on the preliminary parameter analysis, the latent dimension is set to $d_z = 16$ and $\beta = 0.5$. The previous tuning results showed that a latent dimension around 16 provides a favorable balance between representation capacity and generalization, while $\beta = 0.5$ offers a suitable compromise between reconstruction quality and latent-space regularization.

The VAE is trained for 40 epochs using the Adam optimizer with a learning rate of 1×10^{-3} and a mini-batch size of 128.

For a test sample x, the latent mean is used to generate a deterministic reconstruction $\hat{x}$. The reconstruction error is defined as

$$e_v(x) = \frac{1}{D} \parallel x - \hat{x} \parallel_2^2 \tag{6}$$

where $D = 784$. A larger reconstruction error generally indicates that the input deviates more strongly from the learned training distribution.

2.4. Quality-Aware Fuzzy Fusion

VAE reconstruction error alone may not completely characterize local image degradation. Therefore, image roughness is introduced as an additional quality indicator. It is calculated from the horizontal and vertical neighboring-pixel variations:

$$r(x) = \frac{1}{N_x}\sum \mid \Delta_x x \mid + \frac{1}{N_y}\sum \mid \Delta_y x \mid \tag{7}$$

The roughness measure is mainly sensitive to local high-frequency changes, while the VAE reconstruction error reflects the difference between the input image and the learned data distribution. The two indicators therefore provide complementary quality information.

To reduce the effect of extreme samples, both indicators are normalized using the median and median absolute deviation (MAD). For a quality variable $v(x)$, the robust normalized score is calculated as

$$z(v) = \frac{v(x) - \text{median}(v_{\text{train}})}{\text{MAD}(v_{\text{train}}) + \varepsilon} \tag{8}$$

The normalized roughness score and VAE reconstruction score are denoted by $q_r(x)$ and $q_v(x)$, respectively. They are combined as

$$q(x) = \lambda q_r(x) + (1 - \lambda) q_v(x) \tag{9}$$

where $\lambda = 0.75$.

A fuzzy degradation membership is then constructed using a sigmoid function:

$$\mu_d(x) = \frac{1}{1 + \exp[-a(q(x) - b)]} \tag{10}$$

where $a = 1.5$ controls the transition steepness and $b = 3.0$ determines the center of the degradation region.

Based on the estimated degradation membership, the contribution of each feature view is dynamically adjusted. The raw-view weight is calculated as

$$w_r(x) = 0.5 + \delta\mu_d(x) \tag{11}$$

where $\delta = 0.48$. The HOG-view weight is therefore

$$w_h(x) = 1 - w_r(x) \tag{12}$$

The final class-probability vector is obtained by

$$\mathrm{p}(x) = w_r(x)\mathrm{p}_r(x) + w_h(x)\mathrm{p}_h(x) \tag{13}$$

When an input image is close to the clean training distribution, $\mu_d(x)$ remains small and the two feature branches contribute approximately equally. As the estimated degradation level increases, $w_r(x)$ gradually increases, while the contribution of the HOG branch is reduced. This allows the model to adapt the fusion strategy to sample quality without introducing a complex gating network.

2.5. Implementation Details

All experiments use a fixed random seed of 42 for the main evaluation. The β-VAE is trained for 40 epochs using Adam with a learning rate of 10^{-3}, a batch size of 128, $\beta = 0.5$, and a latent dimension of 16. The RBF-SVM penalty parameters are set to $C = 10$ for the raw-intensity branch and $C = 3$ for the HOG branch. The fuzzy fusion parameters are fixed as $\lambda = 0.75$, $a = 1.5$, $b = 3.0$, $\delta = 0.48$.

Robustness is further evaluated under three types of image degradation without retraining the classifiers. Gaussian noise is introduced with standard deviations of $\sigma \in \{0.05, 0.10, 0.15, 0.20\}$ while random masking removes 10%, 20% and 30% of image pixels. Gaussian blur is applied using standard deviations of$\sigma_b \in \{0.5, 1.0\}$.

These corruption settings are used to analyze whether the proposed quality-aware fuzzy fusion mechanism can dynamically adjust the contribution of different representations when the input quality changes.

3. Results and Discussion

3.1. Overall Classification Performance

Table I compares the overall classification performance of different methods on the Fashion-MNIST dataset. The proposed FuzzyVAE-QFusion achieves an accuracy of 0.860 and a Macro-F1 score of 0.8556, outperforming VAE-Latent SVM, MLP, Extra Trees, Raw RBF-SVM, and HOG RBF-SVM. As shown in Fig. 2, VAE-Latent SVM gives the lowest classification accuracy of 73.6%, while MLP and Extra Trees achieve 77.4% and 81.7%, respectively. Raw RBF-SVM and HOG RBF-SVM further improve the accuracy to 82.7% and 84.2%. These results indicate that both raw intensity information and structural HOG features are useful for fashion image classification. By adaptively combining the two complementary representations according to image quality, FuzzyVAE-QFusion achieves the best overall classification performance among the compared methods.

Table 1. Overall classification performance of different methods on Fashion-MNIST.

Model	Accuracy	Macro-F1	Macro-Precision	Macro-Recall	NLL
VAE-Latent SVM	0.736	0.733375	0.741531	0.733131	0.735118
MLP	0.774	0.766165	0.7672	0.772789	0.660959
Extra Trees	0.817	0.808721	0.813202	0.814845	0.634726
Raw RBF-SVM	0.827	0.822056	0.825203	0.824903	0.509376
HOG RBF-SVM	0.842	0.838181	0.838874	0.83852	0.459485
FuzzyVAE-QFusion (Proposed)	0.86	0.855582	0.856113	0.857861	0.453375

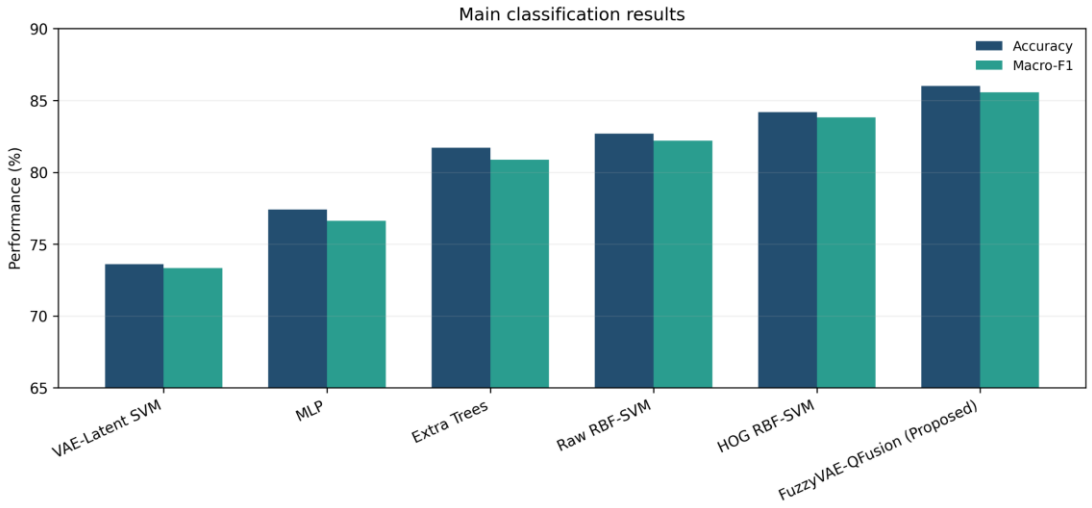

Figure 2. Comparison of classification accuracy and Macro-F1 among different methods.

In addition to accuracy and Macro-F1, FuzzyVAE-QFusion achieves a Macro-Precision of 0.8561 and a Macro-Recall of 0.8579. Its negative log-likelihood (NLL) is 0.4534, which is also lower than those of the compared baseline models. This indicates that the proposed method provides both improved classification performance and reliable probability estimates.

Fig. 3 further shows the class-wise classification performance of FuzzyVAE-QFusion. High recognition rates are obtained for Trouser, Sandal, Bag, and Ankle boot, with normalized recall values of 0.95, 0.93, 0.97, and 0.95, respectively. In contrast, Shirt remains the most difficult category, with a recall of only 0.54. Most Shirt errors occur with T-shirt/top, Pullover, and Coat. This is reasonable because these upper-body clothing categories have similar shapes and local structures. Therefore, the remaining errors are mainly caused by strong visual similarity between categories rather than a general failure of the proposed model.

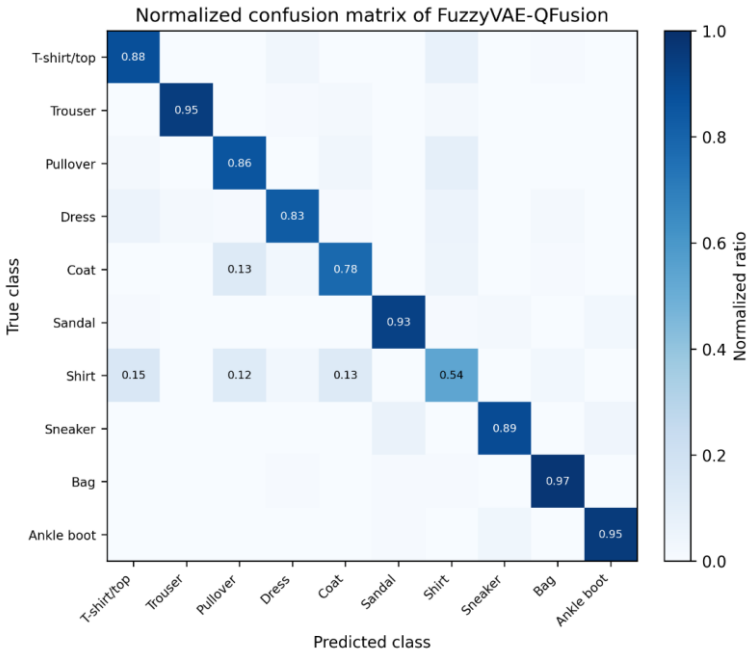

Figure 3. Normalized confusion matrix of FuzzyVAE on the Fashion-MNIST test set.

3.2. Robustness under Image Corruptions

To further evaluate model reliability, Gaussian noise, random masking, and Gaussian blur are introduced into the test images. Figs. 4 and 5 present the results under Gaussian noise and random masking, while all three corruption types are included in the overall robustness evaluation. HOG

RBF-SVM is highly sensitive to noise, and its accuracy decreases rapidly from approximately 84% on clean images to about 15% when the Gaussian noise standard deviation reaches 0.20. This indicates that local gradient-based structural features can be strongly affected by high-frequency noise.

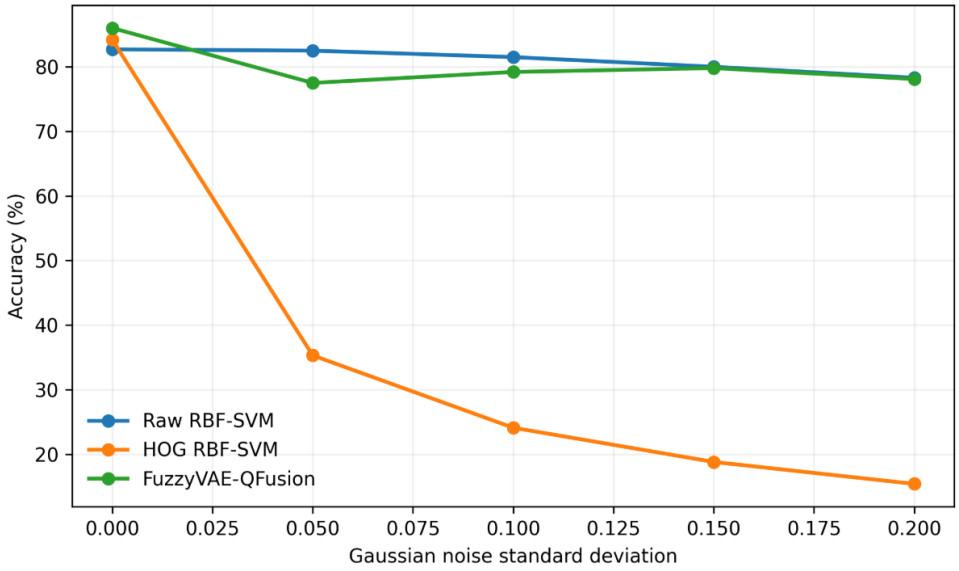

Figure 4. Classification robustness under different levels of Gaussian noise.

In contrast, both Raw RBF-SVM and FuzzyVAE-QFusion show much more stable performance. When the Gaussian noise standard deviation increases to 0.20, their accuracies remain close to 78%. Raw RBF-SVM performs slightly better under several noisy conditions, showing that raw intensity information is relatively stable against Gaussian noise. However, FuzzyVAE-QFusion starts from a higher clean-image accuracy and gradually increases the contribution of the raw-intensity branch as the estimated degradation becomes stronger. This allows the proposed method to maintain a good balance between clean-image classification and noise robustness.

A similar trend can be observed under random masking in Fig. 5. When 30% of the image pixels are masked, the accuracy of HOG RBF-SVM decreases to approximately 46.5%, while FuzzyVAE-QFusion maintains an accuracy of about 66.5%. Raw RBF-SVM achieves a slightly higher accuracy of about 68.4% under this severe masking condition. Nevertheless, the proposed method provides higher clean-image performance while remaining close to Raw RBF-SVM under increasing masking ratios. This result shows that the adaptive fusion strategy can reduce the effect of unreliable structural information when image quality decreases.

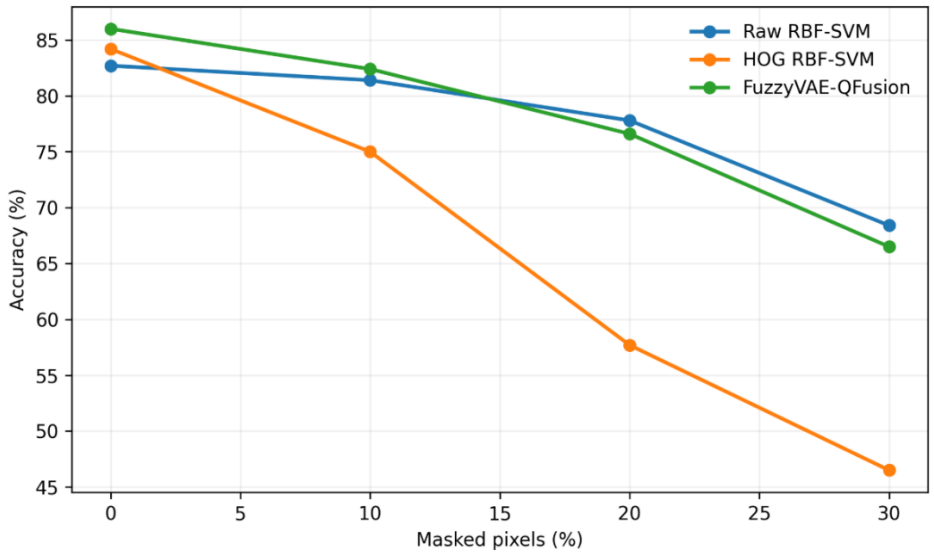

Figure 5. Classification robustness under different random masking ratios.

Table II summarizes the overall robustness results. FuzzyVAE-QFusion achieves a clean accuracy of 0.860 and a mean corruption accuracy of 0.7871, resulting in the highest composite score of 0.8236. Raw RBF-SVM obtains a slightly higher mean corruption accuracy of 0.7927, but its clean accuracy is lower at 0.827, leading to a lower composite score of 0.8098. HOG RBF-SVM achieves a mean corruption accuracy of only 0.4842 because of its high sensitivity to noise and masking. These results show that the main advantage of FuzzyVAE-QFusion is not simply achieving the highest accuracy under every individual corruption condition, but providing a better overall balance between clean-image performance and robustness under image degradation.

Table 2. Clean performance and average robustness under image corruptions.

Model	Clean Accuracy	Mean Corruption Accuracy	Composite Score
Raw RBF-SVM	0.82	0.792666667	0.809833
HOG RBF-SVM	0.84	0.484222222	0.663111
FuzzyVAE-QFusion	0.86	0.787111111	0.823556

3.3. Analysis of the Quality-Aware Fuzzy Gate

Figure 6 shows how the proposed fuzzy gate alters its fusion strategy under different image conditions. For clear images, the average weight of the original view is approximately 0.55, indicating that the contributions of the original image branch and the HOG branch are similar. When the image quality deteriorates, the degradation score increases, and the fuzzy gate will assign a greater weight to the original image branch.

For example, under Gaussian noise with $\sigma = 0.05$, the weight of the original view increases to approximately 0.61; when $\sigma = 0.10$, the weight is about 0.79; when σ reaches 0.15, the weight exceeds 0.94. At $\sigma = 0.20$, the weight is close to 0.98. Similar trends were also observed under random masking conditions. These results are consistent with the robustness results in Figures 4 and 5, that is, as the image damage intensifies, the reliability of HOG features decreases.

For Gaussian blur, an interesting phenomenon was observed: the image quality score decreases, while the weight of the original view remains around 0.50. This indicates that the effect of blur on the two feature representations is different from that of Gaussian noise and random masking. More importantly, the fuzzy gate does not always increase the weight of the original view but adjusts the fusion ratio according to the estimated image condition. This result confirms that FuzzyVAE-QFusion performs sample quality-aware adaptive fusion rather than fixed weighted fusion, which explains its stronger robustness under different image damage conditions.

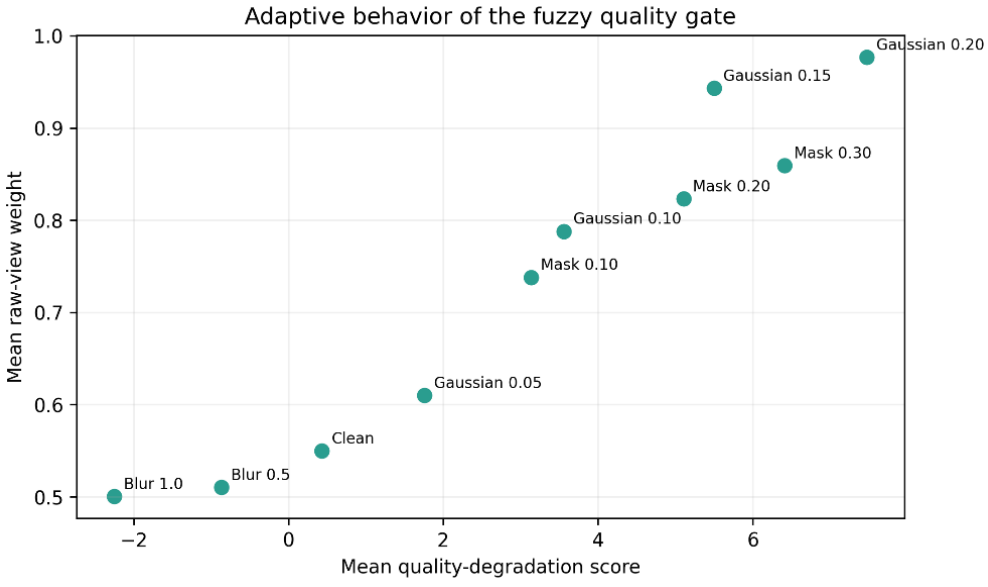

Figure 6. Adaptive behavior of the proposed fuzzy quality gate under different image conditions.

4. Conclusion

When the test images are affected by noise, masking or other quality variations, robust fashion image classification remains a challenge. This paper proposes a quality-aware fuzzy fusion framework named FuzzyVAE for classification on the Fashion-MNIST dataset. This method combines classifiers based on original intensity and HOG, and uses the variational autoencoder (VAE) to estimate the input quality. The reconstruction error of VAE and image roughness are jointly used to measure the degree of sample degradation, and the fuzzy membership function dynamically adjusts the contribution of the two feature views. Different from static fusion, the proposed method adjusts the fusion weights according to the specific situation of each input image. Experimental results show that FuzzyVAE achieves an accuracy of 86.0% and a macro F1 score of 85.56% on clean images. More importantly, in the presence of Gaussian noise, random masking, and blurring, it achieves an average damaged accuracy of 78.71% and a maximum comprehensive score of 82.36%. The fuzzy gate analysis further confirms that as the degree of image degradation increases, the model gradually reduces the contribution of unreliable structural features. These results indicate that FuzzyVAE provides a simple and effective solution for balancing classification performance and robustness in the case of image condition changes.

References

[1] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," arXiv preprint arXiv:1708.07747, 2017.

[2] Y. Guo, D. Stutz, and B. Schiele, "Improving robustness of vision transformers by reducing sensitivity to patch corruptions," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023, pp. 4108–4118.

[3] K. Berahmand, F. Daneshfar, E. S. Salehi, Y. Li, and Y. Xu, "Autoencoders and their applications in machine learning: A survey," Artif. Intell. Rev., vol. 57, Art. no. 28, 2024, doi: 10.1007/s10462-023-10662-6.

[4] A. Solera-Rico, C. Sanmiguel Vila, M. Gómez-López, Y. Wang, A. Almashjary, S. T. M. Dawson, and R. Vinuesa, "-Variational autoencoders and transformers for reduced-order modelling of fluid flows," Nat. Commun., vol. 15, Art. no. 1361, 2024, doi: 10.1038/s41467-024-45578-4.

[5] Q. Yu, S. Dai, R. Dong, and S. Ikuno, "Attention-based vector quantized variational autoencoder for anomaly detection by using orthogonal subspace constraints," Pattern Recognit., vol. 164, Art. no. 111500, 2025, doi: 10.1016/j.patcog.2025.111500.

[6] Y. M. Manu, R. Gagana, and S. V. Shashikala, "Enhancing face detection and recognition through machine learning algorithm," in Proc. 2023 Int. Conf. Recent Advances Sci. Eng. Technol. (ICRASET), 2023, doi: 10.1109/ICRASET59632.2023.10420184.

[7] X. Gu, P. P. Angelov, and Q. Shen, "Semi-supervised fuzzily weighted adaptive boosting for classification," IEEE Trans. Fuzzy Syst., vol. 32, no. 4, pp. 2318–2330, Apr. 2024, doi: 10.1109/TFUZZ.2024.3349637.

[8] S. S. Sumit, S. Anavatti, M. Tahtali, S. Mirjalili, and U. Turhan, "ResNet-Lite: On improving image classification with a lightweight network," Procedia Comput. Sci., vol. 246, pp. 1488–1497, 2024, doi: 10.1016/j.procs.2024.09.597.

[9] Z. Xu, M. A. A. Aziz, and N. A. Razak, "Research on the application of adaptive parameter adjustment strategy in RBF kernel SVM image classification," in Proc. 2024 IEEE 6th Symp. Computers Informatics (ISCI), 2024, pp. 138–145, doi: 10.1109/ISCI62787.2024.10668339.

[10] Z. Xu, M. A. A. Aziz, and N. A. Razak, "Performance evaluation of support vector machine algorithm in object classification using different preprocessing methods," in Proc. 2024 IEEE Int. Conf. Automatic Control and Intelligent Systems (I2CACIS), Shah Alam, Malaysia, 2024, pp. 169–174, doi: 10.1109/I2CACIS61270.2024.10649625.