ARTICLE

A Degradation Type Adaptive and Deep CNN-Based Image Classification Model for Degraded Images

Huanhua Liu, Wei Wang[*], Hanyu Liu, Shuheng Yi, Yonghao Yu and Xunwen Yao

School of Information Technology and Management, Hunan University of Finance and Economics, Changsha, 410205, China

*Corresponding Author: Wei Wang. Email: wangwei@hufe.edu.cn

ABSTRACT

Deep Convolutional Neural Networks (CNNs) have achieved high accuracy in image classification tasks, however, most existing models are trained on high-quality images that are not subject to image degradation. In practice, images are often affected by various types of degradation which can significantly impact the performance of CNNs. In this work, we investigate the influence of image degradation on three typical image classification CNNs and propose a Degradation Type Adaptive Image Classification Model (DTA-ICM) to improve the existing CNNs' classification accuracy on degraded images. The proposed DTA-ICM comprises two key components: a Degradation Type Predictor (DTP) and a Degradation Type Specified Image Classifier (DTS-IC) set, which is trained on existing CNNs for specified types of degradation. The DTP predicts the degradation type of a test image, and the corresponding DTS-IC is then selected to classify the image. We evaluate the performance of both the proposed DTP and the DTA-ICM on the Caltech 101 database. The experimental results demonstrate that the proposed DTP achieves an average accuracy of 99.70%. Moreover, the proposed DTA-ICM, based on AlexNet, VGG19, and ResNet152, exhibits an average accuracy improvement of 20.63%, 18.22%, and 12.9%, respectively, compared with the original CNNs in classifying degraded images. It suggests that the proposed DTA-ICM can effectively improve the classification performance of existing CNNs on degraded images, which has important practical implications.

KEYWORDS

Image recognition; image degradation; machine learning; deep convolutional neural network

1 Introduction

Image classification is a fundamental task in computer vision and artificial intelligence that has a broad range of applications [1,2]. In recent years, the accuracy of image classification algorithms has been greatly improved with the application of deep Convolutional Neural Networks (CNNs) [3,4]. For instance, the model proposed in [5] surpassed human performance on the ImageNet dataset [6] with a top-5 test error of only 4.94%, while ResNet [7] achieved a top-5 accuracy of 97.8%. CNN-based image classification algorithms have become mainstream due to their high accuracy, which is continually being improved through advances in deep learning technology.

The CNNs have demonstrated outstanding performance on high-quality image classification databases such as ImageNet and PASCAL VOCs [8]. However, images obtained in real-world scenarios are susceptible to various types of degradation due to poor lighting conditions, camera instability, limited exposure time, and other factors. Motion blur is a common degradation that can occur in automatic driving due to object movement during exposure and camera instability. Low resolution is also a prevalent degradation in video surveillance since the camera is often far from the objects of interest [9]. To reduce storage and transmission requirements, original images are typically compressed [10,11], producing compression degradation. Additionally, salt-and-pepper noise and Gaussian white noise are easily introduced during image acquisition and transmission due to external environmental effects, such as electromagnetic interference. Fig. 1 illustrates examples of images polluted by various types of degradation. Such degradations not only impact the perceptual quality of the image but also affect the performance of visual algorithms. Research [12] has shown that carefully selected samples with a small amount of degradation can fool even the best-performing CNN models, while study [13] have demonstrated that both artificial and authentic noise can lead to a decrease in CNN model accuracy. In [13], the effects of artificial noise on different CNN architectures were studied on a relative small database. The effects of several types of artificial and authentic noise were investigated in [14]. This study aims to investigate the effects of artificial noise on CNNs using a larger, popular image classification database, meanwhile we proposes a method to improve the accuracy of existing CNNs in classifying degraded images.
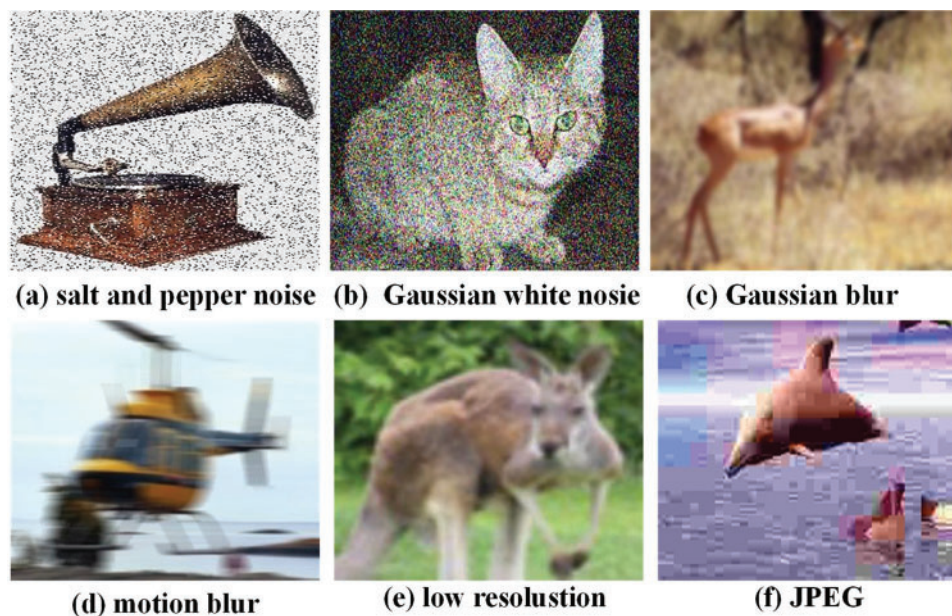

Figure 1: (a)–(f) show the visualization degradation examples of salt and pepper noise, Gaussian white noise, Gaussian blur, motion blur, low resolution, and JPEG compression degradation, respectively

To our knowledge, two main approaches have been proposed to improve existing CNNs for degraded image classification: improving the quality of the degraded image through degradation-specific de-noising algorithms and training the CNNs using noised samples. While pre-processing degraded images with de-noising algorithms can improve both image quality and CNN classification performance [15–20], improvements using CNN-based de-noising are limited [21]. Recent studies [21,22] have shown that training CNNs on noise-contaminated samples with the same type of noise can

improve classification accuracy. Since the number of the common degradation types is limited [23,24], it is feasible to train a Degradation Type Specified Image Classifier (DTS-IC) for each degradation type. Grouping similar degradation types [25] can further reduce the number of required DTS-ICs. Therefore, we propose the Degradation Type Adaptive Image Classification Model (DTA-ICM) for degraded images, which trains existing models on degraded images to build DTS-IC for each degradation type and uses a Degradation Type Predictor (DTP) to predict the degradation type of the degraded image. The main contributions of this work are threefold:

1. Through experimental methods, we analyze the impact of six frequently encountered types of degradation on three prominent CNNs used in image classification.
2. To address the challenge of degraded image classification, we propose the DTA-ICM approach, which predicts the degradation type using the DTP, and subsequently activates the corresponding DTS-IC to facilitate classification.
3. We propose the deep CNN based DTP to predict the degradation type of degraded image, and the proposed DTP and DTA-ICM are evaluated on Caltech 101 database [26], of which the generalization ability is also studied.

The paper is organized as follows. The effects of degradation on CNNs of image classification are analysed in Section 2. The DTA-ICM is proposed in Section 3. The experiments and analysis are shown in Section 4. Section 5 concludes this paper.

2 Degradation Effects of Image Classification CNNs

2.1 Image Degradations

In practical scenarios, images acquired from different sources often suffer from various types of degradation. In this work, we considered six common types of degradation as follows:

- **Salt-and-pepper noise**, a typical impulse noise commonly observed in surveillance camera imaging, is characterized by randomly replacing original pixels with black and white pixels. We synthesized the degradation level of salt-and-pepper noise, denoted as σ_{sp}, which represents the percentage of original pixels replaced by noise.
- **Gaussian white noise** may be introduced if an image is transmitted through a poor channel or captured using a low quality senor. In synthesizing degradation, a noise map with the same spatial dimension with the image was generated firstly, which has zero mean and standard deviation σ_{gw} being used to control the degradation level. Secondly, the RGB channels were added with the same noise.
- **Gaussian blur,** which often occurs in post-processing operations. We generate zero mean Gaussian blur as that in [13], let σ_{gb} denote the square Gaussian kernel, and the standard deviation of the blur was obtained as that in [13].
- **Motion blur** typically arises due to poor camera stabilization or object movement during exposure. We defined σ_{mb} as the blur kernel width, which is the number of pixels contributing to the motion blurring.
- **Low resolution**, another common degradation in surveillance cameras with low quality, was synthesized by down-sampling the high quality reference image with step S in this work.

- **JPEG compression degradation** occurs during image compression using the popular JPEG encoder. Images captured from various sources are often compressed before storage, transmission, or further processing, which leads to compression degradation. Quality Factor (QF) controls the compressed quality in JPEG, where a higher QF denotes better quality.

For the mentioned parameters, a bigger value of σ_{sp}, σ_{gw}, σ_{gb}, σ_{mb}, and S means a heavier degradation. Fig. 2 shows some images degraded from the above six types of degradation with different levels.
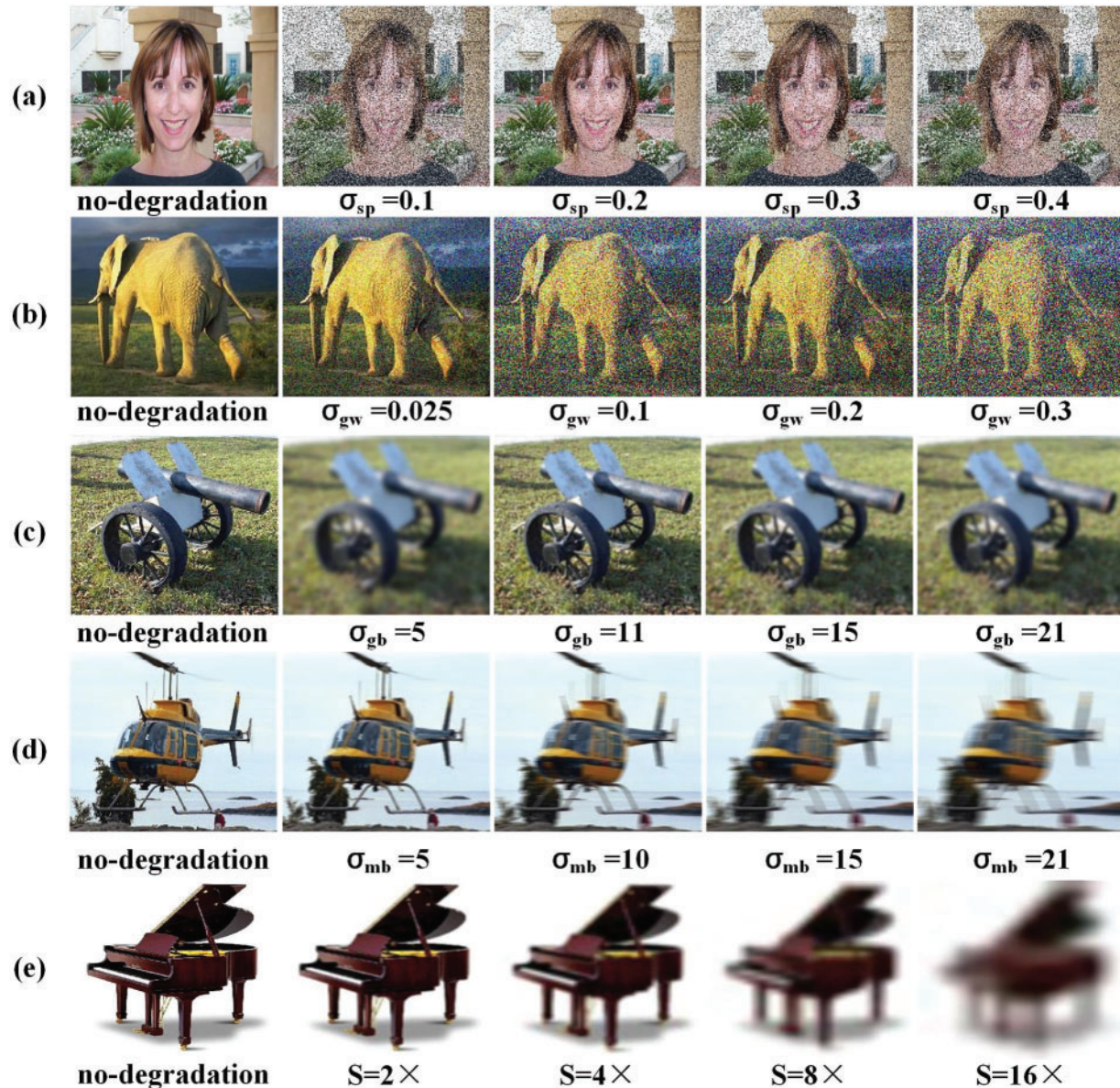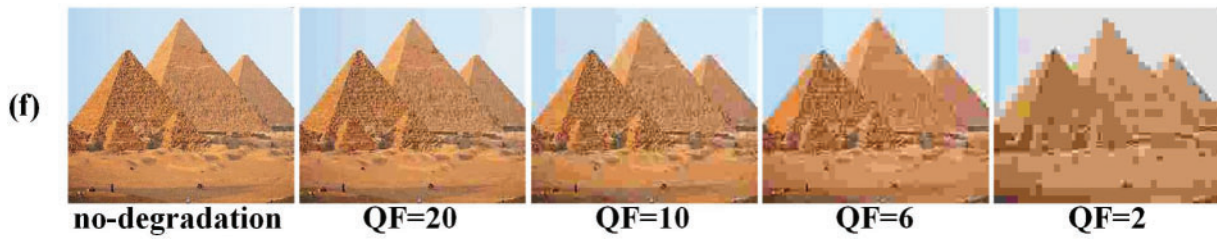

Figure 2: (Continued)

Figure 2: Examples of degraded images with different degradation levels. (a) to (f) are salt-and-pepper noise, Gaussian white noise, Gaussian blur, motion blur, low resolution (down sampling), and JPEG compression, where σ_{sp}, σ_{gw}, σ_{gb}, σ_{mb}, S, and QF are degradation level parameters

2.2 Effects of Image Degradation

To investigate the impact of various degradation types on CNN-based image classifiers, three widely used models, namely AlexNet [3], VGG19 [4], and ResNet152 [7], were chosen and the Caltech 101 [26] dataset was employed as the test dataset. Firstly, the Caltech 101 dataset was randomly split into training, validation, and test subsets with an 8:1:1 ratio, and the selected CNNs were pre-trained on the ImageNet dataset and fine-tuned on the training subset of Caltech 101. Next, the test subset of Caltech 101 was used to synthesize images with six previously mentioned types of degradations, namely salt-and-pepper noise, Gaussian blur, motion blur, low resolution, and JPEG compression degradation. The degradation levels were set to $\sigma_{sp} = 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, ..., 1.0$, $\sigma_{gs} = 0.025, 0.05, ..., 0.275$, $\sigma_{gb} = 5, 11, 15, ..., 51$, $\sigma_{mb} = 5, 11, 15, ..., 51$, $S = 2, 4, 8, 16$, $QF = 60, 50, 40, 30, 20, 10, 8, 6, 4, 2$. Finally, the well-trained models were evaluated on the synthesized images from the Caltech 101 test subset, and the results are shown in Fig. 3, where the x and y axes represent the degradation level index and top1 accuracy, respectively. It is observed that all the models exhibit similar sensitivity to the degradations, resulting in decreased accuracy with increasing degradation levels. This decline in accuracy is attributed to the original training of the models on high-quality images, making them very sensitive to the five types (excluding JPEG) of degradation, where even a small degradation can cause a rapid decrease in accuracy. Additionally, the accuracy also drops when QF is lower than 30 for JPEG degradation. To address these issues and enhance the accuracy of CNN-based classifiers in classifying degraded images, we propose the DTA-ICM, which is discussed in Section 3.
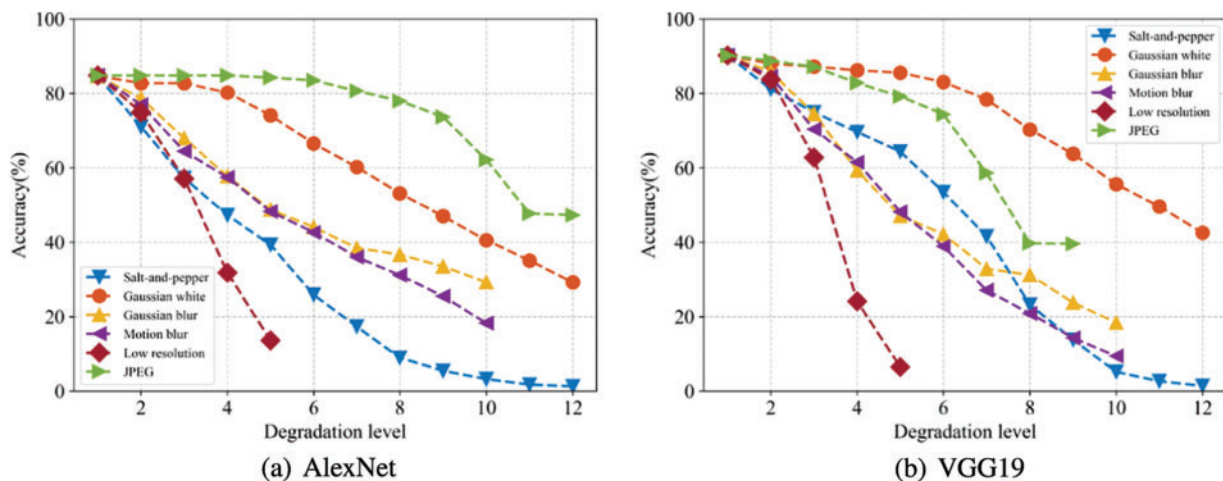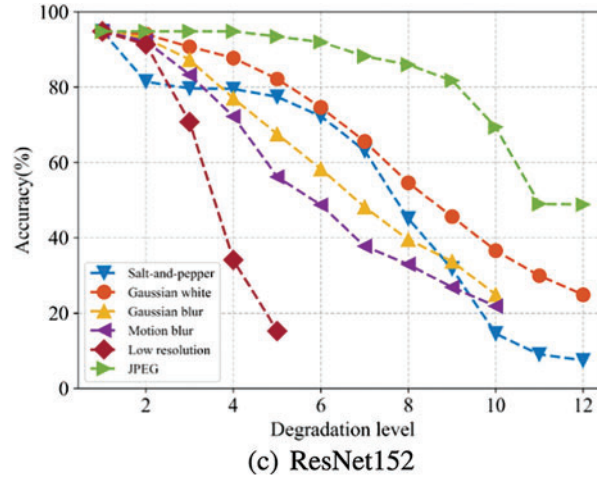

Figure 3: (Continued)

Figure 3: The top1 accuracy of degraded test subset of Caltech 101 dataset. (a) AlexNet. (b) VGG19. (c) ResNet152

3 Proposed Degradation Type Adaptive Image Classification Model

3.1 Framework

Motivated by the observation that a deep CNN classifier trained on the images with the same type of degradation can improve its performance in classifying degraded images, we propose the DTA-ICM for degraded image classification, as shown in Eq. (1).

$$y = F(T(\phi(x), x), \phi(x)), \tag{1}$$

where x represents the degraded image, which may be affected by one or more types of degradation. The function $\phi()$ is a degradation type predictor that predicts the types of degradation that x belongs to. Based on the results of $\phi()$, the corresponding DTS-IC or DTS-ICs $T()$ will be selected from the DTS-IC set (i.e., existing CNNs trained for each degradation type) to classify x. $F()$ is the fusion model that combines the classification results of $T()$ and the degradation type prediction results of $\phi()$ to obtain the final classification result for x, which may use a weighting scheme. Specifically, we mainly consider images with only one type of degradation in this work, meaning that a single type of degradation will be predicted for x, and a specified image classifier for that type of degradation will be activated to classify x.

Fig. 4 shows the framework of the proposed DTA-ICM including input of a degraded image, DTP, image classifier selection strategy, DTS-IC. The DTP aims to predict the possibility of degradation type for the input image as $P = P_0, P_1, P_2, ..., P_N$, where N is the number of degradation types. In this work, we take six common types of degradation mentioned in Section 2.1 and special no-degradation case into consideration, i.e., N is six, P_0 denotes the no-degradation case, and the DTP is a seven-class classifier. Let $C = C_0, C_1, C_2, ..., C_N$ denotes the DTS-IC set, where C_i is the image classifier for the ith type degradation which is obtained by training the existing CNN image classifier on the images with i^{th} type degradation. In this work, AlexNet, VGG19, ResNet152 are choosen as the existing CNNs, the number of degradation types considered in this work is six, denoted by N. Specifically, C_0 represents no degradation, i.e., the original CNNs. The number of DTS-ICs is equal to N, and an increase in the number of degradation types leads to an increase in the complexity of the DTA-ICM. In the case where the number of degradation types is large, clustering similar degradation types can be employed

to reduce the number of DTS-ICs. For instance, motion blur and Gaussian blur can be grouped into the blur degradation type, where a single image classifier for the blur degradation group is trained to recognize both motion blur and Gaussian blur. The image classifier selection strategy aims to select the appropriate DTS-IC for image classification based on the prediction result P. In this work, images affected by a single degradation type are considered. The ith type of degradation is predicted if P_i has the maximum value, and C_i is then selected for classification, of which the results are the final prediction results of the proposed DTA-ICM in this work.
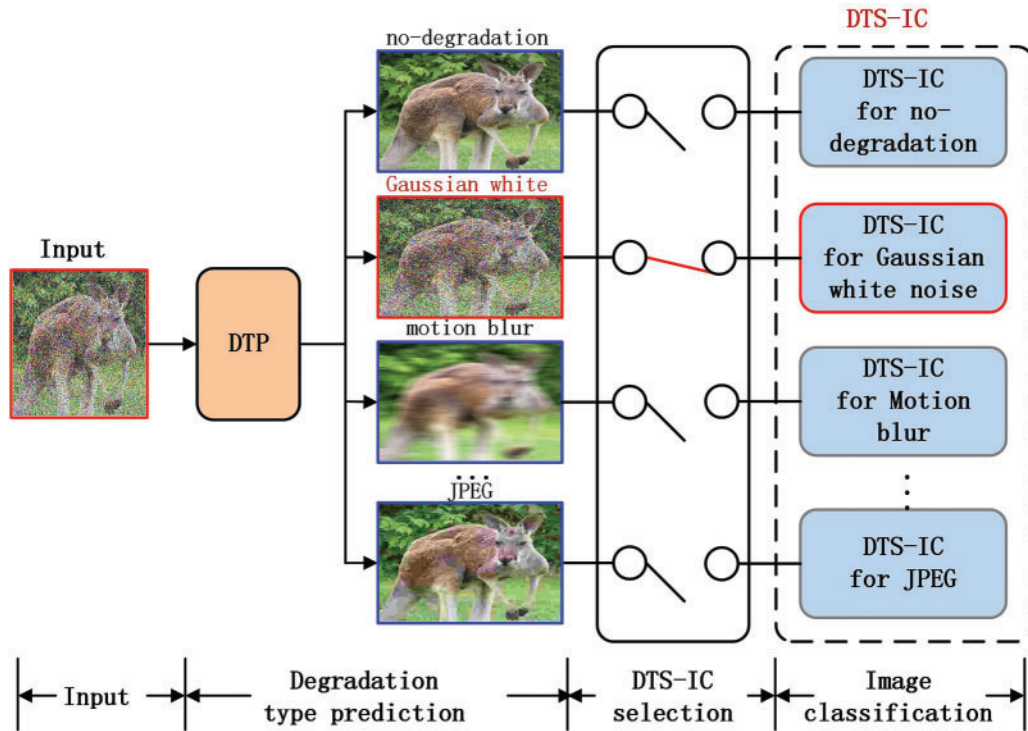

Figure 4: Framework of the proposed degradation type adaptive image classification model

3.2 Degradation Type Prediction Model

The task of predicting the degradation type of a degraded image x can be formulated as a multi-class classification problem represented by the equation

$$z = \Psi(g(x)), \tag{2}$$

where $g()$ denotes the feature extraction model for image, $\Psi()$ is a multi-class classification function, which can be the Softmax function. In this study, we consider seven classes, comprising six common degradation types, namely salt-and-pepper noise, Gaussian white noise, Gaussian blur, motion blur, low resolution, and JPEG compression, as well as the no-degradation type. The human visual system has a hierarchical structure where low-level semantic features are extracted in the early visual region, while complex higher-level semantic features are extracted in the higher visual region. Based on this structure, CNNs have been developed to simulate the image feature extraction process of the human visual system, the low-level network extracts low-level image features such as edge and brightness, and the high-level network realizes the extraction of high-level image features such as image patterns and content. These networks have achieved remarkable success in visual tasks such as image classification, object detection, and target tracking. With the increase of the depth of the CNN, the network will

has a more powerful feature representation ability and the accuracy of classification will increase. The residual network (ResNet) has been widely used in various computer vision tasks, since it is a very deep CNN that overcomes the problems of gradient explosion and disappearance by incorporating "shortcut connection". Therefore, the network of the proposed DTP in this work is mainly based on ResNet50, which is shown Fig. 5. It consists of eight layers, the first layer consists of a convolutional layer with kernel size 7×7 and step 2, and a maximum pooling layer. From *Conv_2×* to *Conv_5×* are five down-sampling residual modules, which consist of 3, 4, 6, and 3 residual blocks including convolutional block and identity block, respectively. Compared with image recognition, degradation type prediction needs more low level semantic features. In order to use the low level feature in prediction, the feature map extracted from the convolutional block of *Conv_3×* (28×28) is down-sampled 4 times by using max pooling, which is then contacted with the high level semantic features from *Conv_5×*. The followed three layers are average pooling, full connected layer of 7 dimension, and SoftMax layer, which finish the degradation type prediction based on the fused feature maps.
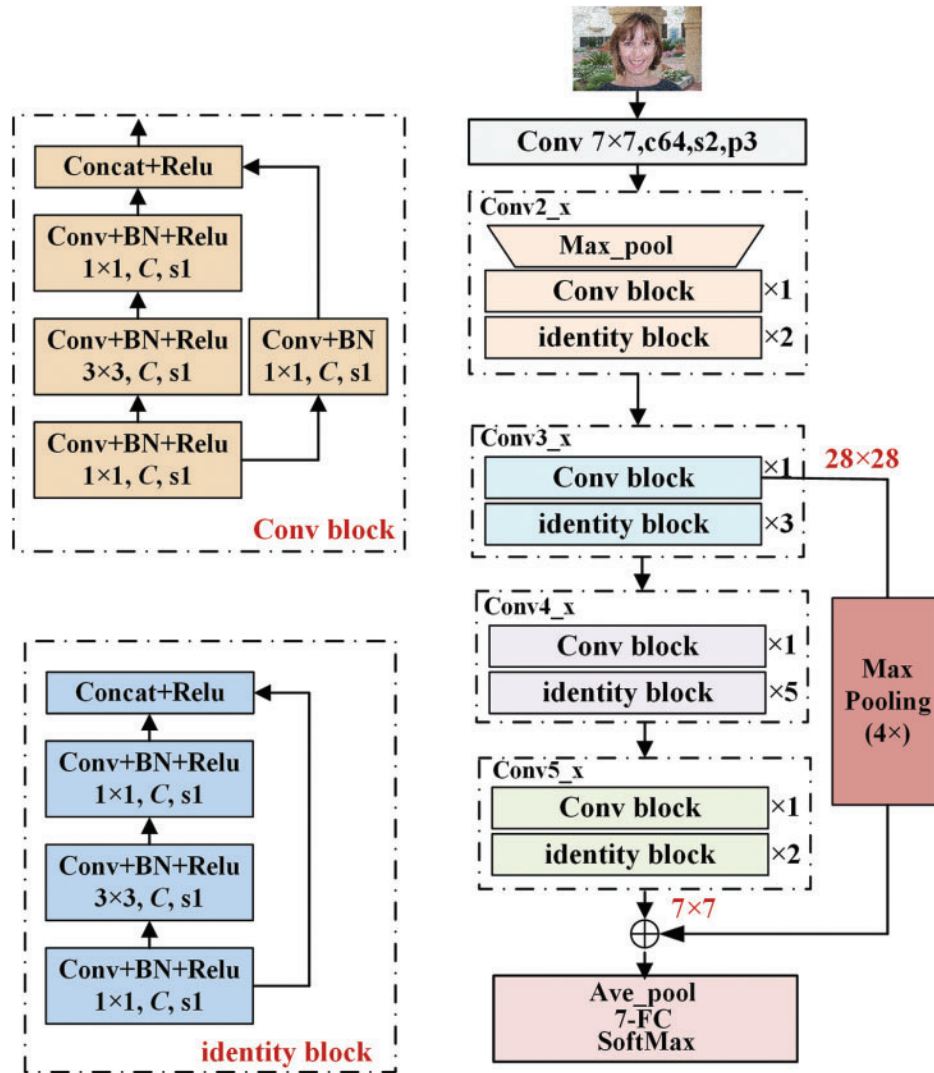

Figure 5: Network of the degradation type prediction model

The degradation type predictor (DTP) in this work is modeled as a multi-class classification problem, for which cross-entropy is used as the loss function, as shown in Eq. (3).

$$L = -\sum_{i=0}^{K-1} y_i log(p_i),$$

(3)

where y_i represents the ground truth label, i.e., degradation type, p_i is the output of the degradation type predictor, and K denotes the number of classes that the proposed DTP, i.e., the number of degradation types considered in this work. Since six common degradation types and no-degradation type are considered in this work, K is seven.

3.3 Degradation Type Specified Image Classifier

AlexNet, VGG19, and ResNet152 are the most widely used CNNs for image classification, demonstrating excellent performance. Therefore, we have chosen these three CNNs as the existing image classifiers. To obtain the DTS-IC set, we initialized the three CNNs on the ImageNet database and the training sub-set of caltech 101 database. Subsequently, we constructed the degradation training data set by synthesizing each of the six types of degradations listed earlier on the training sub-set images of caltech 101. Finally, we re-trained each CNN using the degraded images synthesized from the caltech 101 training sub-set, resulting in the DTS-IC set.

4 Experimental Results and Analysis

4.1 Setup

The experiments were conducted on a graphics workstation equipped with an Intel Core I9-10900X CPU, 64 GB RAM, and an Nvidia GeForce RTX 3080 Ti GPU with 12 GB memory. We employed the Caltech 101 image classification database to train and evaluate the proposed model. The database consists of 101 categories and each category contains images ranging from 40 to 800. The Caltech 101 dataset was randomly divided into training, validation, and test subsets in an 8:1:1 ratio. To simulate degradation, we applied salt-and-pepper noise, Gaussian white noise, Gaussian blur, motion blur, low-resolution, and JPEG compression with parameters $\sigma_{sp} = 0.075$, $\sigma_{gw} = 0.2$, $\sigma_{gb} = 15$, $\sigma_{mbw} = 15$, $S = 4$, $QF = 8$, respectively. We used the prediction accuracy as the evaluation metric, defined as the number of samples with all correct classifications divided by the total sample size, which is given by the equation

$$Accuracy = \frac{TP + TN}{P + N},$$

(4)

where TP represents the number of positive samples classified correctly, TN is the number of negative samples classified incorrectly, P and N are the total positive and negative samples, respectively.

In order to make the image size of Caltech 101 to be consistent with the input size of DTP, cropping and interpolation were adopted. Since the intensity of natural images varies greatly, we locally normalize the images to enhance the robustness of the model to changes of intensity and contrast. Firstly, the RGB color map is converted to gray image, and the local contrast normalization is realized by Eq. (5).

$$\hat{I} = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + c},$$

(5)

where $I(i,j)$ is the intensity value at the position (i,j), c is a small positive constant, $\mu(i,j)$ and $\sigma(i,j)$ represent the mean and variance shown as Eqs. (6) and (7), respectively,

$$\mu(i,j) = \sum_{p=-W}^{p=W} \sum_{q=-H}^{q=H} I(i+p, j+q), \tag{6}$$

$$\sigma(i,j) = \sqrt{\sum_{p=-W}^{p=W} \sum_{q=-H}^{q=H} (I(i+p, j+q) - \mu(i,j))^2}, \tag{7}$$

where W and H are the width and height of the local window, respectively.

4.2 Evaluation of the Proposed Degradation Type Predictor

To evaluate the performance of the proposed DTP, we selected AlexNet, VGG19, and ResNet50 as the comparison models. The backbone of these models was used as a feature extractor, and the fully connected layer was modified as a seven-class classification layer. All models were trained on the ImageNet database and then fine-tuned on the Caltech 101 training subset with synthesized degradations. The accuracy results are shown in Table 1. The average accuracy of ResNet50 and the proposed DTP were 97.95% and 99.70%, respectively, which are higher than that of AlexNet (92.85%) and VGG19 (95.21%). The DTP showed a slight improvement compared to ResNet50, which can be attributed to the introduced fusion model that fuses low features together with high features. From the perspective of degradation types, AlexNet and VGG19 exhibited poor performance on Gaussian white noise and no-degradation cases, with accuracies of 78.86% and 89.21%, respectively. The no-degradation accuracy of ResNet50 was 94.36%, which is lower than that of the proposed DTP (99.27%). The proposed DTP demonstrated excellent performance for all degradation types, with accuracies of over 99% for most degradations, except for motion blur, which had an accuracy of 98.82%. Thus, we can conclude that the proposed DTP achieved excellent performance for both the average accuracy and each degradation type.

Table 1: Accuracy of the proposed DTP (%)

	Salt-and-pepper	Gauss. white	Gauss. blur	Motion blur	Low resolution	JPEG	No-degrad.	Ave.
AlexNet	98.82	78.86	96.53	94.65	95.21	96.72	89.21	92.85
VGG19	99.91	79.53	**100**	**99.94**	99.40	97.21	90.36	95.21
ResNet152	99.94	94.53	**100**	99.12	99.3	98.40	94.36	97.95
Proposed	**100**	**100**	**100**	98.82	**100**	**99.84**	**99.27**	**99.70**

In this section, we present an evaluation of the generalization ability of the proposed DTP to different degradation levels. During training, we chose only one degradation level for each degradation type, which were set to $\sigma_{sp} = 0.075$, $\sigma_{gw} = 0.2$, $\sigma_{gb} = 15$, $\sigma_{mb} = 15$, $S = 4$, QF $= 8$ for salt-and-pepper noise, Gaussian white noise, Gaussian blur, motion blur, low-resolution, and JPEG, respectively. For testing, we varied the degradation levels to $\sigma_{sp} = 0.025$, 0.05, ..., 0.1, 0.2, ..., 1.0, $\sigma_{gw} = 0.075$, 0.1, ..., 0.3, 0.35, 0.4, $\sigma_{gb} = 5$, 10, ..., 40, 50, $\sigma_{mb} = 5$, 10, ..., 40, 50, $S = 2$, 4, 8, QF $= 0$, 2, 4, ..., 10, 20, respectively, and the results are shown in Fig. 6. For salt-and-pepper noise, the proposed DTP's performance was not affected by the variation in degradation levels. The accuracy of the proposed DTP remained stable for the other five degradation types at most degradation levels. However, the accuracy decreased rapidly when the degradation is too small, specifically for the leftmost five points, where the degradation parameter were $\sigma_{gw} = 0.075$, $\sigma_{gb} = 5$, $\sigma_{mb} = 5$, $S = 2$, QF $= 20$, respectively. Due

to the limited degradations, some images were misclassified as no-degradation, and these images were classified by the original classifier in the proposed DTA-ICM. The accuracy of misclassified samples was 79.82%, 88.91%, and 92.73% for AlexNet, VGG19, and ResNet152, respectively, which is close to that of the original CNNs for no-degraded images. This result indicates that the proposed DTA-ICM also worked well in such cases. For JPEG compression, the accuracy also decreased rapidly when QF = 2 due to severe degradation. However, in practical applications, QF = 2 is unlikely to be used for image compression due to its poor compression quality. Therefore, we conclude that the proposed DTP has a robust performance across different degradation levels for all degradation types.
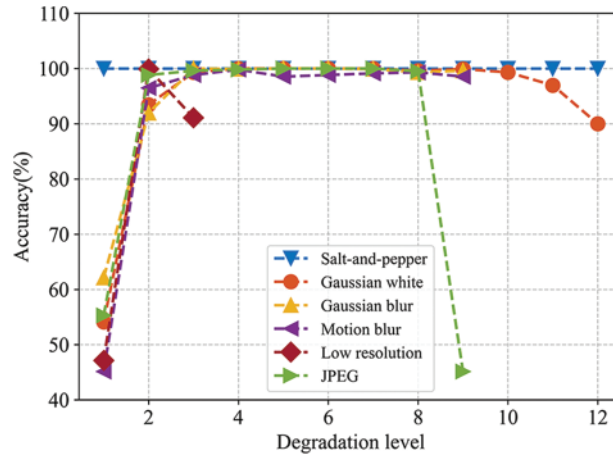

Figure 6: The accuracy of the proposed DTP on different degradation levels

4.3 Evaluation of the Proposed Degradation Type Adaptive Image Classification Model

We evaluated the effectiveness of the proposed DTA-ICM by comparing its performance with that of the original AlexNet, VGG19, and ResNet152 models. For this purpose, all models were fine-tuned and tested on the Caltech 101 dataset with synthesized degradations. In our proposed model, the DTP was first employed to predict the degradation type of the degraded images, which were then classified by the DTS-IC. The results of our experiments, as shown in Table 2, demonstrate that the proposed model achieved higher average accuracies of 81.97%, 87.74%, and 92.21% based on AlexNet, VGG19, and ResNet152, respectively, compared to the original CNNs, which achieved accuracies of only 61.34%, 69.52%, and 79.31%, respectively. The proposed model also showed significant improvements in accuracy for each degradation type, such as 13.1%, 16.03%, 20.49%, 13.98%, and 20.60% for ResNet152 in salt-and-pepper, Gaussian white noise, Gaussian blur, low resolution, and JPEG compression, respectively. Our results suggest that the proposed DTA-ICM can effectively improve the accuracy of existing CNNs for degraded image classification.

**Table 2:** The accuracy of proposed DTA-ICM (%)

		Salt-and-pepper	Gauss. white	Gauss. blur	Motion blur	Low resolution	JPEG	No-degrad.	Ave.
AlexNet	Org. [3]	47.33	47.05	57.49	57.78	57.09	77.98	84.68	61.34
	DTA-ICM	**80.35**	**81.86**	**82.46**	**83.01**	**80.08**	**81.32**	**84.68**	**81.97**

(Continued)

Table 2 (continued)

		Salt-and-pepper	Gauss. white	Gauss. blur	Motion blur	Low resolution	JPEG	No-degrad.	Ave.
VGG19	Org. [4]	69.64	63.79	61.58	59.31	62.83	79.23	90.24	69.52
	DTA-ICM	**87.02**	**89.13**	**87.63**	**86.21**	**86.38**	**87.54**	**90.24**	**87.74**
ResNet152	Org. [7]	79.63	74.63	72.25	77.07	70.77	85.98	94.84	79.31
	DTA-ICM	**92.73**	**90.66**	**92.74**	**91.05**	**91.37**	**92.11**	**94.84**	**92.21**

The proposed DTA-ICM comprises the DTP and DTS-IC components, as illustrated in Section 1. The DTP demonstrates high accuracy, achieving an average accuracy of 99.7%, indicating that it can effectively predict the six degradation and non-degradation types. In this section, we evaluate the generalization ability of the DTA-ICM on different degradation levels by assessing the DTS-IC's generalization ability in Gaussian blur. We trained the DTS-IC of AlexNet, VGG19, and ResNet152 with a degradation level parameter of $\sigma_{gb} = 15$ and tested with σ_{gb} values of 0, 5, 11, 15, 21, 25, and 31. The results shown in Fig. 7 indicate that all DTS-IC models achieve the highest accuracy when the degradation level in testing matches that in training. The accuracy decreases with increasing differences between the training and testing degradation levels. Nonetheless, the accuracy remained stable and high between $\sigma_{gb} = 11$ and $\sigma_{gb} = 25$, indicating the proposed model's excellent generalization ability.
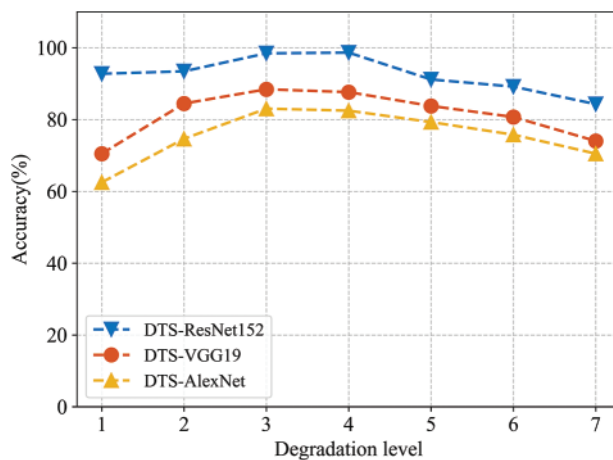
Figure 7: The accuracy of the proposed DTS-IC on different degradation levels

5 Conclusions

The current Convolutional Neural Network (CNN)-based image classification models are usually trained on no-degraded images, which can lead to decreased accuracy in real-world scenarios where images can be degraded. This work investigated the impact of six common types of degradation on the accuracy of three typical CNN models using experimental methods. The results demonstrate that all types of degradation can significantly reduce the accuracy of CNNs. To address this issue, we proposed a Degradation Type Adaptive Image Classification Model (DTA-ICM), comprising a Degradation Type Predictor (DTP) and a set of Degradation Type Specified Image Classifiers

(DTS-IC). The experimental results show that the proposed DTP has an average accuracy of 99.70%, with accuracy exceeding 98% for each degradation type, indicating its effectiveness in predicting degradation type. Compared to the original CNNs of image classification, the proposed DTA-ICM based on AlexNet, VGG19, and ResNet152 models yields an accuracy increase of 20.63%, 18.22%, and 12.9%, respectively, demonstrating the efficacy of the proposed model in enhancing existing CNN-based image classification models in degraded image classification.

Author Contributions: Study conception and design: H. Liu, W. Wang; data collection: H. Liu, S. Yi, Y. Yu, X. Yao; analysis and interpretation of results: H. Liu, H. Liu, S. Yi, Y. Yu, X. Yao; draft manuscript preparation: H. Liu, W. Wang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Yang, J., Yu, K., Gong, Y., Huang, T. (2009). Linear spatial pyramid matching using sparse coding for image classification. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1794–1801. Miami, FL, USA.
2. Sanchez, J., Perronnin, F. (2011). High-dimensional signature compression for large-scale image classification. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1665–1672. Colorado Springs, CO, USA.
3. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60(6),* 84–90. https://doi.org/10.1145/3065386
4. Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
5. He, K., Zhang, X., Ren, S., Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. *IEEE International Conference on Computer Vision*, pp. 1026–1034. Washington, DC, USA.
6. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K. et al. (2009). ImageNet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. Miami, FL, USA.
7. He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778. Las Vegas, NV, USA.
8. Everingham, M., van Gool, L., Williams, C. K., Winn, J., Zisserman, A. (2010). The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision, 88(2),* 303–338.

9.   Zou, W. W., Yuen, P. C. (2012). Very low resolution face recognition problem. *IEEE Transactions on Image Processing, 21(1),* 327–340.

10. Liu, H., Zhang, Y., Zhang, H., Fan, C., Sam, K. et al. (2020). Deep learning-based picture-wise just noticeable distortion prediction model for image compression. *IEEE Transactions on Image Processing, 29(1),* 641–656.

11. Zhang, Y., Liu, H., Yang, Y., Sam, K., Kuo, C. C. et al. (2022). Deep learning based just noticeable difference and perceptual quality prediction models for compressed video. *IEEE Transactions on Circuits and Systems for Video Technology, 32(3),* 1197– 1212.

12. Moosavi-Dezfooli, S. M., Fawzi, A., Frossard, P. (2016). Deepfool: A simple and accurate method to fool deep neural networks. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2574–2582. Las Vegas, NV, USA.

13. Roy, P., Ghosh, S., Bhattacharya, S., Pal, U. (2018). Effects of degradations on deep neural network architectures. arXiv preprint arXiv:1807.10108v4.

14. Pei, Y., Huang, Y., Zou, Q., Zhang, X., Wang, S. (2022). Effects of image degradation and degradation removal to CNN-based image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(4),* 1239–1253.

15. Li, R., Pan, J., Li, Z., Tang, J. (2018). Single image dehazing via conditional generative adversarial network. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8202–8211. Salt Lake City, UT, USA.

16. Qu, Y., Chen, Y., Huang, J., Xie, Y. (2019). Enhanced pix2pix dehazing network. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8160–8168. Long Beach, CA, USA.

17. Lu, B., Chen, J. C., Chellappa, R. (2019). Unsupervised domainspecific deblurring via disentangled representations. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10225–10234. Long Beach, CA, USA.

18. Baek, I., Davies, A., Geng, Y., Rajkumar, R. R. (2018). Realtime detection, tracking, and classification of moving and stationary objects using multiple fisheye images. *IEEE Intelligent Vehicles Symposium*, pp. 447–452. Changshu, China.

19. Moller, T., Nillsen, I., Nattkemper, T. W. (2017). Active learning for the classification of species in underwater images from a fixed observatory. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2891–2897. Venice, Italy.

20. Dong, C., Loy, C. C., He, K., Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2),* 295–307.

21. Liu, H., Yu, Y., Liu, S., Wang, W. (2022). A military object setection model of UAV reconnaissance image and feature visualization. *Applied Science, 12(23)*. 12236. https://doi.org/10.3390/app122312236

22. Dodge, S., Karam, L. (2016). Understanding how image quality affects deep neural networks. *IEEE International Conference on Quality of Multimedia Experience*, pp. 1–6. Lisbon, Portugal.

23. Wu, J., Ma, J., Liang, F., Dong, W., Shi, G. et al. (2020). End-to-end blind image quality prediction with cascaded deep neural network. *IEEE Transactions on Image Processing, 29,* 7414–7426.

24. Lin, H., Hosu, V., Saupe, D. (2019). KADID-10k: A large-scale artificially distorted IQA database. *Eleventh International Conference on Quality of Multimedia Experience*, pp. 1–3. Berlin, Germany.

25. Fan, C., Zhang, Y., Feng, L., Jiang, Q. (2018). No reference image quality assessment based on multi-expert convolutional neural networks. *IEEE Access, 6,* 8934–8943.

26. Li, F., Fergus, R., Perona, P. (2006). One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(4),* 594–611. https://doi.org/10.1109/TPAMI.2006.79