Fidelity Estimation Improves Noisy-Image Classification with Pretrained Networks

Xiaoyu Lin, Deblina Bhattacharjee, Majed El Helou, Members, IEEE, and Sabine Süssstrunk, Fellow, IEEE.

Abstract—Image classification has significantly improved using deep learning. This is mainly due to convolutional neural networks (CNNs) that are capable of learning rich feature extractors from large datasets. However, most deep learning classification methods are trained on clean images and are not robust when handling noisy ones, even if a restoration preprocessing step is applied. While novel methods address this problem, they rely on modified feature extractors and thus necessitate retraining. We instead propose a method that can be applied on a pretrained classifier. Our method exploits a fidelity map estimate that is fused into the internal representations of the feature extractor, thereby guiding the attention of the network and making it more robust to noisy data. We improve the noisy-image classification (NIC) results by significantly large margins, especially at high noise levels, and come close to the fully retrained approaches. Furthermore, as proof of concept, we show that when using our oracle fidelity map we even outperform the fully retrained methods, whether trained on noisy or restored images.

Index Terms—Noisy-Image Classification, Deep Learning, Image Restoration, Data Fidelity

I. INTRODUCTION

Traditionally, image classification methods have relied on a limited set of hand-crafted features [1]–[6]. Deep learning approaches, especially supervised methods using convolutional neural networks (CNNs) [7], [8] trained on large annotated datasets [9], are capable of extracting rich feature representations [9], [10] and thus improve performance. However, obtaining reliable and generalizable feature extractors remains a challenge [11]–[14]. It is thus desirable to exploit existing, pretrained feature extractors in a modular way across datasets or applications without the need to retrain them. Here, we exploit the feature extractor pretrained on clean images to work with noisy input.

Classification methods trained on datasets with clean images show a significant performance drop on noisy or restored images [15]. Yet, in many applications, the captured images are inherently noisy, for instance, in low-light conditions, short exposure capture, in multi-spectral acquisition, and when employing inexpensive surveillance cameras, etc. A recent study [16] confirmed that image classification accuracy drops significantly when the input images are noisy. Furthermore, even a straight-forward restoration preprocessing step failed to improve classification.

Recently, two studies specifically addressed Noisy-Image Classification (NIC). However, they either require costly retraining [17] or even necessitate an entire redesign of the CNN-based classifier [18].

In contrast, our proposed approach can be extended to any feature extractor without retraining, enabling existing networks to also classify noisy images. Our Fidelity-Guided Noisy-Image Classification (FG-NIC) method exploits an internal integration of a data fidelity estimator within the feature extractor to adjust the features at test time. We estimate the fidelity map with a deep network. We then exploit the fidelity map to guide the attention of the pretrained feature extractor by fusing its weights into the different network layers (see Fig. 1). Experimental evaluations show that our method improves the baseline by large margins and even rivals fully retrained classifiers. Furthermore, our proof-of-concept method with an oracle fidelity map outperforms even all the fully retrained classification networks.

II. RELATED WORK

Deep networks outperform human classification on clean images [19]. However, their performance deteriorates on noisy images on which humans do better [19]. A recent study [16] demonstrates the negative effect of degraded images, including hazy images, underwater images, and motion-blurred images, on classification accuracy. Their results parallel those of an earlier work [20]. Both studies emphasize the need to develop robust NIC methods. Validating these observations, two recent approaches, DeepCorrect [17] and WaveCNets [18], address the problem of NIC under additive white Gaussian noise (AWGN). The authors of DeepCorrect [17] analyze how...
feature extractor activations are affected by noise. To do so, DeepCorrect estimates how susceptible each filter is to the AWGN degradation and learns to replace these filters by retraining them. WaveCNets proposes a wavelet-decomposition network architecture [18]. As noise mostly affects the high-frequency components of an image, WaveCNets trains a more noise-robust network that focuses on low frequencies. Both these methods need retraining and thus do not benefit from pretrained classifiers and feature extractors.

The most recent NIC method proposes an ensemble network consisting of a restoration network followed by two classification ones for clean and for degraded images [21]. Though requiring retraining, this method does exploit one pretrained network for some clean images. To be modular and be able to deploy novel pretrained classifiers despite noise, our solution exploits only pretrained classifiers. This way, we avoid the need for computationally expensive retraining, and can directly make use of future, potentially improved, classifiers.

Another way to approach NIC is to apply a restoration preprocessing step. However, when preprocessing with a restoration network, for instance the DnCNN [22] denoiser, the performance of the classification network is still severely affected, which we show in our experimental section. There is an observable trade-off between the perceptual quality of the restoration output and the classification performance that was recently highlighted in [23]. For that reason, we rely on a fidelity map that we can either learn explicitly or, as we show in our supplementary material, learn internally in an end-to-end approach [24]. We use this data fidelity to guide the attention of the network [25], [26] and account for potential distortions coming from the degradation or from the restoration steps. Learning an internal indicator for various degradation levels was recently shown to improve the performance and robustness of deep denoisers [27]. In our case, we use fidelity as an attention guide for adjusting the activations of the feature extractor in a classifier.

III. METHOD

We present our proposed FG-NIC method by first introducing our fidelity map that we fuse into the pretrained classifier. We discuss the high-level manipulation of the pretrained classifier and present the details of our fidelity fusion. The overall pipeline is shown in Fig. 1. An extended figure with the specific fidelity fusion details is in the supplementary material.

A. Fidelity map

A fundamental element of our approach is the fidelity map that we use to adjust the feature extractor’s output and to guide the attention of the network. Given an image restored by a pretrained network, some image components are more faithful to the original clean image than others. This can depend on the noise (possibly spatially varying), the frequency components or the texture across the image, and on the pretrained restoration network itself. We define the fidelity map in terms of the element-wise $\ell_1$ distance between the original clean image $X$ and the restored image $\tilde{X} = R(X)$, where $R(\cdot)$ is the restoration network and $\tilde{X}$ is the noisy image corresponding to $X$. In the supplementary material, we further test fidelity maps based on $\ell_2$ and cosine distances. While the $\ell_2$ distance performs comparably to the $\ell_1$ distance for lower noise levels, the $\ell_1$ distance outperforms it at higher noise levels, due to its robustness to outliers. We hence rely on a fidelity map based on the $\ell_1$ distance. This fidelity map at pixel $(i,j)$ is

$$F_{(i,j)} = 1 - \frac{1}{C} \sum_{c=1}^{C} |\tilde{X}_{(i,j,c)} - X_{(i,j,c)}|$$

where $\tilde{X}_{(i,j,c)}$ is the pixel intensity of the restored image at pixel $(i,j)$ and channel $c$, and similarly $X_{(i,j,c)}$ is the pixel intensity of the clean image $X$, and $C$ is the number of channels (3 for RGB). All images and the fidelity map are defined over $[0,1]$. From our definition for the fidelity map, it indicates our confidence level in each restored pixel.

The ground-truth oracle fidelity map is naturally not available at test time. Therefore, we train a deep network to act as the fidelity map estimator for the deep restoration network that we use. In our experiments, we implement the fidelity estimator using a state-of-the-art residual architecture [22].

B. Manipulated classifier

For a given CNN-based image classification network $f(\cdot)$, we refer to its feature extractor part that outputs the high-level features for an image as $\phi(\cdot)$. And we refer to its feature classifier, which mainly consists of fully connected layers for predicting the class likelihoods, as $\psi(\cdot)$. We formulate our method as a manipulation over the pretrained feature extractor $\phi(\cdot)$. As we discuss in detail in the following section, we exploit our fidelity map to guide the attention of the pretrained feature extractor by fusing its information into the different network layers. This enables us to adapt the importance that the network assigns to different features based on the varying quality of the degraded-then-restored image. The manipulated classification network can then be formulated generally as $y = \psi(\phi'(\tilde{X}))$ where $\phi'(\tilde{X})$ is the output of the manipulated feature extractor, relative to the original extractor $\phi(X)$, and $\psi(\cdot)$ are the fully connected layers that output the class likelihoods $y$. Our approach enables us to rely on $\phi(\cdot)$ and $\psi(\cdot)$ taken directly from a classification network that is pretrained on clean images. Hence, without any retraining, we can integrate a novel classification network in a straight-forward manner by obtaining $\phi'(\cdot)$ as described in the following section.

C. Fidelity fusion

To obtain the manipulated feature extractor, $\phi'(\cdot)$, relative to the original pretrained extractor $\phi(\cdot)$, we build our fusion over $\phi(\cdot)$ and describe our Algorithm 1 in what follows.

Spatial linear fusion. The input restored image $\tilde{X}$ is affected by the degradation and restoration process. Even when noise is spatially uniform, different spatial regions in an image are affected to varying degrees based on their frequency content [28]. The objective of our fidelity map is thus to guide the attention of the feature extractor accordingly. To that
Algorithm 1 FG-NIC: Fidelity fusion.\(^1\)

Input: \(\tilde{X}\) # Input restored image
Input: \(F\) # Our fidelity map
Input: \(\phi() = \text{Layer}^\phi(\cdot)(\cdot)\) # Pretrained extractor
Input: [\(\text{Block}_1^\phi(\cdot), \text{Block}_2^\phi(\cdot), \ldots\)] # CNN for multiplication
Input: [\(\text{Block}_1^f(\cdot), \text{Block}_2^f(\cdot), \ldots\)] # CNN for addition
Input: \(\text{Block}_1^c(\cdot)\) # FC block for channel fusion
Input: \(\text{Block}_c(\cdot)\) # FC block for channel concatenation

# Spatial linear fusion
1: \(x_m \leftarrow \text{Block}_m^0(F); x_c \leftarrow \text{Block}_c^0(F); t \leftarrow 1\)
2: \(x \leftarrow (2 \times \text{Sigmoid}(x_m) \otimes \tilde{X}) \oplus x_c\)
3: for \(n = 1, \ldots, N\) do
4: if \(\text{Layer}^\phi(\cdot) \in \text{PoolingLayers}\) then
5: \(x_m \leftarrow \text{DownSampling}(\text{Block}_m^i(x_m) + f_m)\)
6: \(x_c \leftarrow \text{DownSampling}(\text{Block}_c^i(x_c) + f_c)\)
7: \(x \leftarrow (2 \times \text{Sigmoid}(x_m) \otimes x) \oplus x_c\)
8: \(t \leftarrow t + 1\)
9: end if
10: end for
11: # Channel fusion
12: \(f_c \leftarrow \text{DownSampling}(\text{Flatten}(F))\)
13: \(x \leftarrow 2 \times \text{Sigmoid}(\text{Block}_f(\cdot)) \odot x\)
14: \(x \leftarrow \text{Block}_c(\text{Concatenate}(f_c, x))\)

Output: Modified feature \(x = \phi'(\tilde{X})\)

---

1\(^{\phi}\) is the Hadamard product, \(\otimes\) the Hadamard product spatially with broadcasting to all channels, and \(\oplus\) the element-wise addition in spatial dimensions with channel broadcasting. FC refers to fully-connected layers.

pretrained network, the resulting feature set retains the same size as the original one.

Ensembling. We also investigate using an ensemble of \(\{\phi(\cdot), \phi'(\cdot)\}\) rather than only our manipulated features. Our ensemble module assigns element-wise weights to the two feature layers to combine them. This approach enables our method to exploit the best of the two feature spaces, without having to modify the pretrained fully connected layers. As we show in our experimental results, preserving the original features can be beneficial at low degradation levels.

**IV. EXPERIMENTS**

A. Data processing

We use all categories of the Caltech-256 dataset [7], and perform the standard processing and augmentation used in the state-of-the-art methods [16], [30]. For our AWGN degradation, the additive noise is independent across color channels. For different degradation levels, we change the standard deviation (\(\sigma\)) of the Gaussian noise. For further evaluation, we implement spatially varying degradation with 1D and 2D variations. The noise level linearly changes with the number of rows or columns for 1D varying noise, and for 2D varying noise, the noise level linearly changes with respect to the Euclidean distance from a random point on the image. Fig. 2 illustrates some examples with different AWGN degradations on an image along with their corresponding fidelity maps.

B. Baseline methods

Aside from the pretrained baseline, we compare against multiple fully pretrained networks. Each of our setups is then tested on images with and without restoration preprocessing. In the Pretrained baseline, the classification network is only trained on clean images, and we do not modify it. We also retrain the classification network on a mixture of noisy and clean images (Retrain on noisy), i.e., \(\sigma \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}\). For the Retrain on restored baseline, the training set consists of the outputs of the restoration network. The restored images are obtained by processing the same images as in the previous setup with the DnCNN denoiser [22].

---

\(^{\phi}\)Pretrained models: https://github.com/LiQiufu/WaveCNet. Note, we do not report the AlexNet performance on the WaveCNet model as the authors do not provide the pretrained model.

---
Our proposed FG-NIC method, as presented in Sec. III, exploits a fidelity map to manipulate the image features obtained by a pretrained classification network. The pretrained classifier is the same network used in the first Pretrained baseline, along with a pretrained preprocessing denoiser [22]. We use a pretrained fidelity estimator, which we obtain with the same training setup as that of the denoiser [22]. In our FG-NIC method we also design a proof-of-concept configuration that uses an oracle fidelity map. This is the ground-truth fidelity map, computed following Eq. 1, that we use instead of an estimated fidelity map. This is to study the performance of our proposed approach independent of the quality of the fidelity map estimator. As an ablation study, we allow the network to fine-tune its own learned fidelity estimation with end-to-end training, and present these extended results in the supplementary material. We also provide in the supplementary material, the number of operations and parameters used by the different approaches.

### C. Experimental results

We compare the accuracy of the state-of-the-art methods DeepCorrect [17] and WaveCNets [18] with our FG-NIC on noisy images. The results in Table I are averaged over all uniform noise levels (0.1 to 0.5). We evaluate the models with AlexNet, ResNet-18, and ResNet-50. We report the performance of pretrained models, pretrained models with restoration preprocessing, retrained models, WaveCNets, and DeepCorrect. We observe that our FG-NIC (pretrained), outperforms all other models, and comes close to the one retrained on noisy data. Our FG-NIC (oracle), which is the proof-of-concept method, consistently reports the best performance.

In Table II, we show classification accuracy on various levels of image noise, both uniform and spatially varying. Our methods significantly outperform the pretrained baseline, whether the latter is tested on noisy or restored images, consistently over all noise levels. Comparing with the fully retrained networks, we note that the accuracy of our method comes close, even though the gap increases at high noise levels. Again, our proof-of-concept method with the ground-truth oracle fidelity map outperforms even the fully retrained networks across all noise levels. We also observe that although our ensemble module can slightly boost the classification performance at low noise levels, it quickly becomes detrimental to the results at higher levels. This indicates that our fidelity-based manipulated features $\phi(X)$ are better than the original ones when noise is present. Interestingly, the retrained baselines also tend to overfit their training distribution and are consequently not robust. For instance, the network retrained on noisy images performs significantly worse when given restored images, and the network retrained on restored images performs significantly worse when given noisy images.

## V. Conclusion

We propose the first method that manipulates the layers of a feature extractor to enable the classification of noisy images by a pretrained classifier. By internally fusing data fidelity information, our FG-NIC model guides the attention of the classifier to improve its robustness to noise. We thus avoid retraining a classification network and still significantly improve over the baseline. With our oracle proof-of-concept approach we even consistently outperform, over all noise levels, the fully retrained classifier that is retrained specifically on the outputs of the restoration network. Our method can be readily extended to different degradation types in future work. It is also modular and can be used with future deep classifiers or restoration networks without the retraining overhead.
Acknowledgement: This work was supported in part by the Swiss National Science Foundation via the Sinergia grant number CRSII5–180359.

REFERENCES


Fidelity Estimation Improves Noisy-Image Classification with Pretrained Networks

Xiaoyu Lin, Deblina Bhattacharjee, Majed El Helou, Members, IEEE, and Sabine Süsstrunk, Fellow, IEEE.

Fig. 3. Detailed module architecture (without ensemble) of our proposed method based on ResNet-50 [1]. FC, Conv and ReLU indicate one fully connected layer, one convolutional layer and a rectified linear unit activation, respectively. The pretrained module (feature extractor) is shown in green, and the modules containing trainable parameters are shown in blue.

TABLE III

<table>
<thead>
<tr>
<th>Methods</th>
<th>Uniform degradation (σ)</th>
<th>Varying</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Pretrained</td>
<td>Test on noisy</td>
<td>83.87</td>
</tr>
<tr>
<td></td>
<td>Test on restored</td>
<td>92.03</td>
</tr>
<tr>
<td>Retrain on</td>
<td>Test on noisy</td>
<td>89.62</td>
</tr>
<tr>
<td>noise</td>
<td>Test on restored</td>
<td>89.85</td>
</tr>
<tr>
<td>Retrain on</td>
<td>Test on restored</td>
<td>84.18</td>
</tr>
<tr>
<td>restored</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FG-NIC (Pretrained) Single</td>
<td>89.85</td>
<td>87.76</td>
</tr>
<tr>
<td>FG-NIC (Oracle)  Single</td>
<td>91.97</td>
<td>90.29</td>
</tr>
</tbody>
</table>

I. EXTENDED EXPERIMENTAL RESULTS

The Caltech-101 [2] dataset is similar to Caltech-256 with fewer images and categories. We select 30 images per class as the training set and keep the same procedure as for Caltech-256. The results are consistent and given in Table III.

II. COMPUTATIONAL COMPLEXITY

We show in Table IV the number of Multiply–accumulate operation (MAC) and trainable parameters used in each network or module, highlighting the computational efficiency of our proposed approach.

TABLE III

<table>
<thead>
<tr>
<th>Methods</th>
<th>Uniform degradation (σ)</th>
<th>Varying</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Pretrained</td>
<td>Test on noisy</td>
<td>83.87</td>
</tr>
<tr>
<td></td>
<td>Test on restored</td>
<td>92.03</td>
</tr>
<tr>
<td>Retrain on</td>
<td>Test on noisy</td>
<td>89.62</td>
</tr>
<tr>
<td>noise</td>
<td>Test on restored</td>
<td>89.85</td>
</tr>
<tr>
<td>Retrain on</td>
<td>Test on restored</td>
<td>84.18</td>
</tr>
<tr>
<td>restored</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FG-NIC (Pretrained) Single</td>
<td>89.85</td>
<td>87.76</td>
</tr>
<tr>
<td>FG-NIC (Oracle)  Single</td>
<td>91.97</td>
<td>90.29</td>
</tr>
</tbody>
</table>

III. ABLATION STUDY

We run a series of experiments on Caltech-256 and conduct an in-depth analysis of the results to show the improvement of each module in our proposed method. For the end-to-end setup, we train our FG-NIC and fidelity map estimator together. For the different fidelity map inputs and outputs, downsampling methods, and generally the ablation studies, we use the oracle setup with the single model. The results are given in Table V and support the statements and design decisions we make in our main manuscript.

TABLE IV

<table>
<thead>
<tr>
<th>Methods</th>
<th>#MACs (G)</th>
<th>#Params (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet-50 (Classification)</td>
<td>4.11</td>
<td>24.03</td>
</tr>
<tr>
<td>Ensemble module</td>
<td>4.11</td>
<td>0.21</td>
</tr>
<tr>
<td>Our FG-NIC</td>
<td>0.08</td>
<td>10.49</td>
</tr>
</tbody>
</table>

REFERENCES
