

# Fine-grained Recognition in the Noisy Wild: Sensitivity Analysis of Convolutional Neural Networks Approaches

Erik Rodner<sup>1</sup>  
erik.rodner@uni-jena.de  
Marcel Simon<sup>1</sup>  
marcel.simon@uni-jena.de  
Robert B. Fisher<sup>2</sup>  
rbf@inf.ed.ac.uk  
Joachim Denzler<sup>1</sup>  
joachim.denzler@uni-jena.de

<sup>1</sup> Computer Vision Group  
Friedrich Schiller University Jena  
Germany  
[www.inf-cv.uni-jena.de](http://www.inf-cv.uni-jena.de)  
<sup>2</sup> University of Edinburgh  
United Kingdom

**Abstract** In this paper, we study the sensitivity of CNN outputs with respect to image transformations and noise in the area of fine-grained recognition. In particular, we answer the following questions (1) how sensitive are CNNs with respect to image transformations encountered during wild image capture?; (2) can we increase the robustness of CNNs with respect to image degradations? and (3) how can we predict CNN sensitivity? To answer the first question, we provide an extensive empirical sensitivity analysis of commonly used CNN architectures (AlexNet, VGG19, GoogleNet) across various types of image degradations. This allows for predicting CNN performance for new domains comprised by images of lower quality or captured from a different viewpoint. We also show how the sensitivity of CNN outputs can be predicted for single images. Furthermore, we demonstrate that input layer dropout or pre-filtering during test time only reduces CNN sensitivity for high levels of degradation.

**How sensitive are CNN approaches?** We analyze the sensitivity of three state-of-the-art CNN architectures, which are widely used in recent works: AlexNet [1], VGG19 [2] and GoogLeNet [3]. We perturb test images of different datasets with common noise types (Figure 2) including Gaussian and pepper noise as well as random color shifts and image transformations like random translations, rotations and flips. Our experiments show the weaknesses of a network that is trained on images which contain almost no noise. This is particularly important in real-world applications, where either low budget cameras are used while the training images are noise free or the lighting conditions changed after training.

**Can we make CNNs more robust?** Our experiments show that even small random noise can lead to a dramatic performance decrease. Now the question naturally arises whether it is possible to increase the robustness either during testing or by adapting the learning. In our paper, we analyze two intuitive and simple ideas for increasing robustness, namely data augmentation by applying input dropout to the training data and image pre-processing using a Gaussian or morphological filter.

**Can we predict CNN sensitivity for a test image?** After the empirical analysis, the question remains whether we can quickly detect images with unstable CNN outputs. This question goes beyond a pure sensitivity study but asks for uncertainty estimates often available for Bayesian methods but not for CNNs. We present an approach (Figure 1) for estimating the sensitivity given an input using a first-order approximation of the output change. Our approach is generic enough to cover many different types of noise. At the same time, this approximation allows us to calculate the sensitivity using the back-propagation algorithm, which is used for training CNN models and hence is already available in most frameworks.

**Summary** The experiments show that the influence especially of common intensity noise is severe even at low noise levels. The reason is a domain shift between noise-free training and pertubated test data. From our study, we can draw several conclusions:

1. The training images should have the same noise level as the test images and care has to be taken even for small noise applied to intensities.
2. Data augmentation during training is not the solution as it decreases the accuracy on noise-free images dramatically and is only beneficial for high noise levels as shown.
3. Noise sensitivity depends on the CNN architecture and VGG19 has shown to be the most robust one.
4. Sensitivity of CNN outputs can be predicted for small noise levels with our technique allowing for uncertainty estimates of CNN outputs.

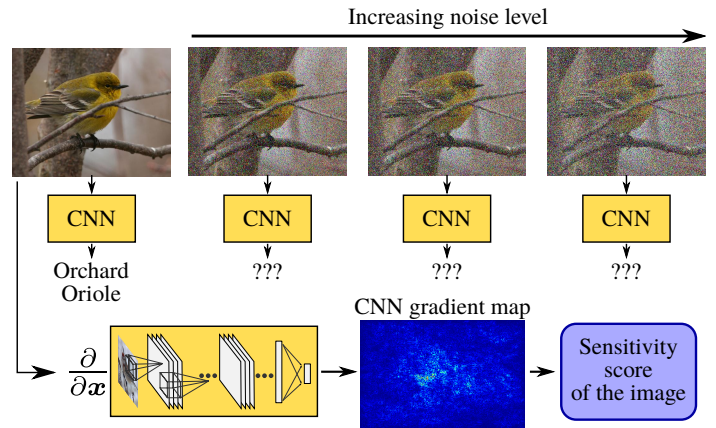


Figure 1: How sensitive are CNNs with respect to image noise and transformations? We study this question and show how to predict CNN sensitivity for a given image.

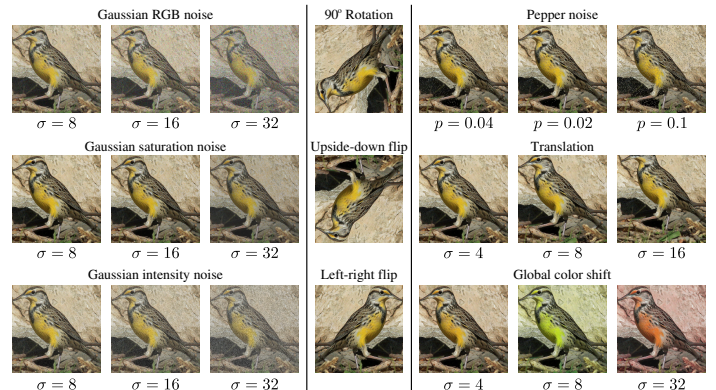


Figure 2: Examples for the image degradations we use in our paper applied to a single image of the CUB-200-2011 dataset [4]. This figure is best viewed in color.

These conclusions can be seen as guidelines especially for developers of real-world applications, where, for example, cheap camera sensors deliver low quality images but the training was performed on relatively noise-free datasets like ImageNet.

- [1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [2] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [3] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*, 2014.
- [4] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.