On Improving Generalization of CNN-Based Image Classification with Delineation Maps Using the CORF Push-Pull Inhibition Operator

Guru Swaroop Bennabhaktula(✉), Joey Antonisse, and George Azzopardi

University of Groningen, Groningen, The Netherlands
g.s.bennabhaktula@rug.nl

Abstract. Deployed image classification pipelines are typically dependent on the images captured in real-world environments. This means that images might be affected by different sources of perturbations (e.g. sensor noise in low-light environments). The main challenge arises by the fact that image quality directly impacts the reliability and consistency of classification tasks. This challenge has, hence, attracted wide interest within the computer vision communities. We propose a transformation step that attempts to enhance the generalization ability of CNN models in the presence of unseen noise in the test set. Concretely, the delineation maps of given images are determined using the CORF push-pull inhibition operator. Such an operation transforms an input image into a space that is more robust to noise before being processed by a CNN. We evaluated our approach on the Fashion MNIST data set with an AlexNet model. It turned out that the proposed *CORF-augmented* pipeline achieved comparable results on noise-free images to those of a conventional AlexNet classification model without CORF delineation maps, but it consistently achieved significantly superior performance on test images perturbed with different levels of Gaussian and uniform noise.

Keywords: CORF · Push-pull · Inhibition · Robustness · Perturbations · Noise suppression · CNN

1 Introduction

In most real-world image classification tasks, there is no control over the environment within which the images are captured. This means that such images might be affected by different types and severity of perturbations (e.g. sensor noise in low-light environments), which may differ from what was present in the training data. Noise is often dynamic and can change over time. Depending on the conditions, noise can suddenly increase due to events in the visual field, which may lead to perturbations in the image affecting the image quality. Examples include adversarial attacks that with very subtle changes to the input images may, for instance, confuse neural networks to classify a panda as a gibbon [8].

N. Tsapatsoulis et al. (Eds.): CAIP 2021, LNCS 13052, pp. 434–444, 2021.
https://doi.org/10.1007/978-3-030-89128-2_42

Image quality directly impacts the reliability and consistency of classification tasks [15]. This challenge has attracted wide interest within the image processing and computer vision communities [5]. Image quality can be affected by a host of factors, such as image compression, during encoding and decoding of images into different formats, resizing, and recoloring, among others. Such methods can also be used as an attack to fool the trained classifier [18]. A common approach to make models more robust to such attacks involves data augmentation during model learning. While data augmentation is effective, its robustness becomes limited when the trained models are deployed into environments where the test images contain noise different than what was present during training.

We hypothesize that giving more importance to the global perceptual contours of a scene will contribute to an image classification solution that is more robust to different types of image noise. To test this hypothesis we use the CORF contour delineation operator with push-pull inhibition, which has been shown to effectively suppress texture and high-frequency noise while delineating the salient contours [3]. We evaluate this transformation tool with respect to different levels of additive perturbations on the Fashion MNIST data set [24] when coupled with the breakthrough network AlexNet [13].

The details of the proposed transformation are presented in Sect. 3.2. Here, we compare two pipelines; a) one that uses the original and noise-free images for training; and b) one that first processes the images with the CORF operator before being fed to the CNN. In order to mimic the real-world scenario where noisy images can be given at the time of model deployment, we evaluate the two pipelines with images consisting of different types and severity of additive noise.

The rest of the paper is organized as follows. In Sect. 2 we present the related works followed by our proposed method in Sect. 3. Experiments and results are reported in Sect. 4, and in Sect. 5 we discuss certain aspects of our work. Finally, we draw our conclusions in Sect. 6.

2 Related Works

Dodge and Karma [7] analyzed how image quality affects the performance of state-of-the-art deep learning models. They trained a network on noise-free images to classify noisy, blurred, and compressed images. From their results, they concluded that image classification is directly proportional to image quality.

In machine/deep learning, this problem can be viewed from the distributions of the training and test data. Ideally, the distributions of the training and test data must be similar for a fair evaluation of the models. In practice, however, the distribution of the test data often deviates from that of the training. In order to account for this unpredictability and make the models more robust, augmented versions of the input data are added to the training set. Data augmentation is among the several techniques used to enhance generalization. Some other popularly used techniques are dropout [20], parameter weight regularization [17], and batch normalization [12], among others. While these techniques are effective, they may not be able to handle deviation in the test set distribution caused due to noise. In order to address this limitation, we propose a transformation step that attempts to enhance the generalization ability of the CNN models in the presence of unseen noise in the test set.

Our hypothesis states that training a model with contour maps of the salient objects instead of the original content results in a classification model that is more robust to unseen noise. This hypothesis requires a robust contour delineation operator that suppresses image noise as much as possible. For this purpose, we use the CORF (Combination of Receptive Fields) operator with push-pull inhibition [3]. It is inspired by the early stages of the mammalian visual system [11], and consists of a system of difference-of-Gaussians (DoG) operators with linearly aligned center-surround areas of support. The output of the operator is an AND-type aggregation of the involved DoG responses. This arrangement is based on the speculation of Hubel and Wiesel [11] that an orientation-selective simple cell is activated when all the afferent LGN cells with center-surround receptive fields are triggered. By means of experiments, it was demonstrated that the CORF model shares more properties with simple cells than the Gabor function model [1]. It is also more effective in contour detection. This operator has later been augmented with two types of inhibition phenomena, namely push-pull [3,21] and surround suppression [14]. It turns out that such inhibition is very effective in suppressing image noise, essentially random strokes and texture that do not belong to the perceptual objects in a given scene.

3 Methods

3.1 Overview

The overall idea is to transform the given images with the CORF contour operator before classification by a CNN model as depicted in Fig. 1.

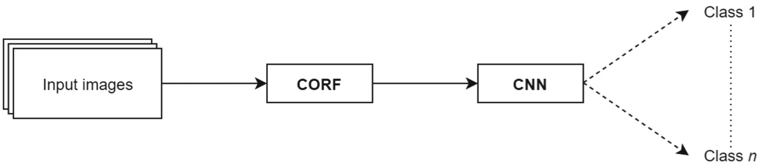


Fig. 1. The proposed application pipeline.

We evaluate the impact on the generalization that the CORF contour operator has on the concerned classification model. Therefore, we compare two pipelines, namely *CORF-free* and *CORF-augmented*. The former is the conventional pipeline that uses the given images as input to the CNN. The latter first delineates the salient contours from the given images by the CORF operator and then uses the resulting contour maps as input to the CNN. Figure 2 illustrates the training and test pipelines of the two approaches.

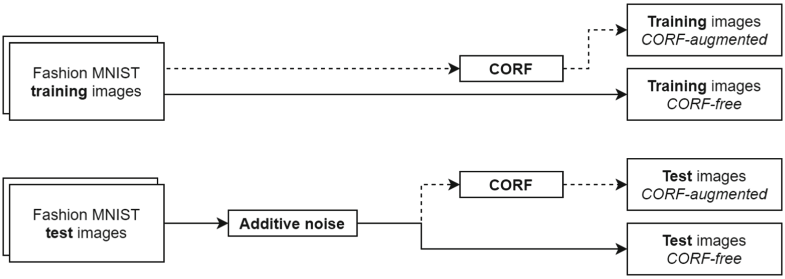


Fig. 2. (Top) Training and (bottom) test pipelines. The solid and dashed arrows indicate the *CORF-free* and the *CORF-augmented* approaches, respectively.

3.2 CORF Operator with Push-Pull Inhibition

The CORF operator is a computational model of orientation-selective simple cells of the mammalian brain [1]. In comparison to the linear Gabor function model, CORF is nonlinear and it achieves more properties of real simple cells; contrast invariant orientation tuning and cross-orientation suppression [3]. The nonlinearity and these two properties result in a CORF operator that is more effective in contour detection than the Gabor function model. The configuration of a model is trainable and its implementation has been found effective in other computer vision [2,9] and signal processing [16] applications.

Figure 3 depicts the structure of a CORF model that is selective for horizontal edges. The circles represent center-on and center-off DoG functions whose output is combined by geometric mean. The standard deviations of the DoG functions and the spacing between their areas of support are hyperparameters of the CORF model used to tune its selectivity. A CORF operator selective for a different orientation can be configured by rotating the alignment of the areas of support of the DoG functions. A rotation-tolerant response can then be achieved by taking the maximum response across all CORF operators selective for different orientations.

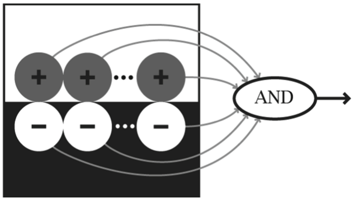


Fig. 3. A CORF computational model of a simple cell that is selective for horizontal edges of the type shown with the white-to-black stimulus behind the circles. The circles indicate the afferent center-on and center-off DoG functions.

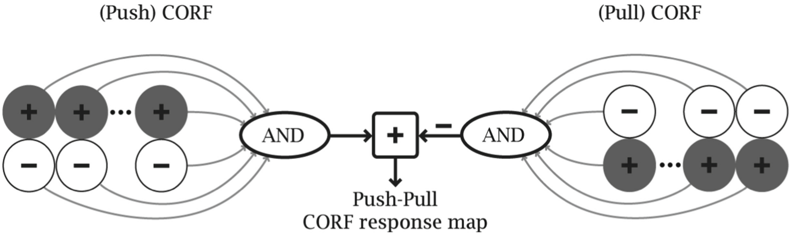


Fig. 4. CORF model with push-pull inhibition. It consists of two sub models, namely push and pull, with the same topology but of opposite selectivity. Their output is then combined with a linear function.

Later, Azzopardi et al. [3] proposed a push-pull CORF model of a simple cell with anti-phase inhibition, which takes as input the response of two CORF models of the type proposed in [1] but with opposing selectivity of luminance contrast. The output of a push-pull CORF model is then the difference between the response of an excitatory (push) CORF model that is stimulated by the pattern of interest and a (weighted) response of the inhibitory (pull) CORF model that is stimulated by the same pattern of interest but of opposite luminance contrast. Figure 4 illustrates the structure of the CORF model augmented with push-pull inhibition. For further technical details, we refer the reader to [3].

In Fig. 5 we illustrate the response maps of the push-pull CORF operator to examples of noise-free and noisy Fashion MNIST images. They demonstrate the operator is very little affected even with high Gaussian noise.

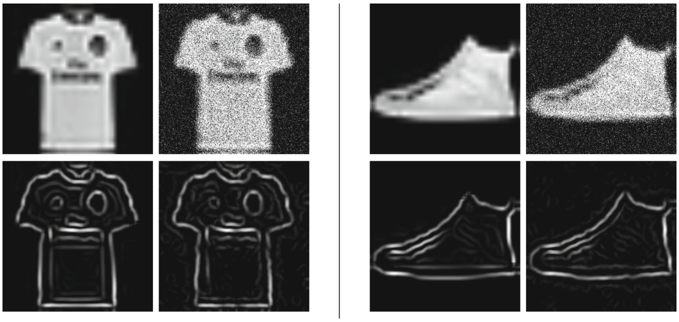


Fig. 5. Robustness of the push-pull CORF delineation operator to Gaussian noise. (Top) Two examples from the Fashion MNIST data set with and without additive Gaussian noise ($\sigma = 0.05$). (Bottom) The corresponding CORF contour maps. The Fashion MNIST images of size 28×28 pixels are resized to 227×227 pixels before the addition of noise.

3.3 AlexNet

We use the AlexNet architecture for our experiments, which was the winning entry in ILSVRC 2012, and was inspired by the Le-Net-5 model introduced in 1998 [26]. AlexNet consists of 8 layers including 5 convolutional layers, 3 fully connected layers, where the final one is the output layer. In order to process grayscale images, the input dimensions of the network are modified to $227 \times 227 \times 1$ pixels from the actual size of $227 \times 227 \times 3$ pixels. The convolutional layers are followed by batch normalization. In our work, batch normalization was used after every convolutional layer which is different from [13], where batch normalization was used only after the first two convolutional layers. The first, second, and the final convolutional layers are followed by a MaxPool layer of size 3×3 pixels with a stride of 2. The first two fully-connected layers consist of 4096 units, each of which is followed by a dropout layer with a factor of 0.5. The number of units in the final fully connected layer is lowered from the original 1000 to 10 classes, the class size of the Fashion MNIST data set. ReLU activations are used in all the convolutional and the fully connected layers, which make training faster in comparison to *tanh* units [13]. The architecture of the AlexNet is depicted in Fig. 6 and for a detailed overview, we refer the reader to [13].

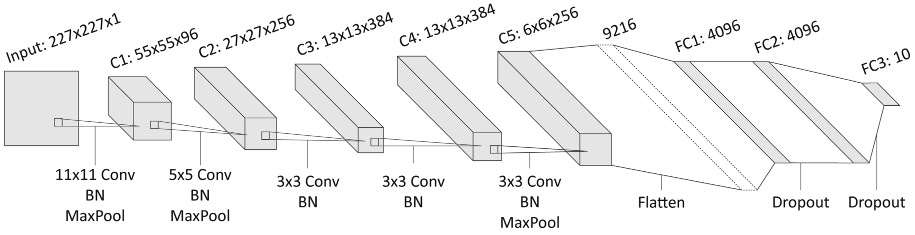


Fig. 6. An illustration of the AlexNet architecture with 5 convolutional (Conv) and 3 fully connected (FC) layers where all the Conv layers are followed by batch normalization (BN).

Sophisticated networks, such as VGG-16 [19] and ResNet-151 [10] have shown to improve the accuracy on ImageNet [6], when compared to AlexNet. However, when it comes to choosing a convolutional network for the relatively simple Fashion MNIST data set [24], we prefer to use AlexNet. This decision is motivated by the fact that the design of AlexNet is simple and it is efficient in terms of time complexity. Due to its simple architectural design and relatively fewer parameters, AlexNet was also found to generalize better when compared to more sophisticated networks [10, 19]. Although we use AlexNet in our experiments, in principle, the proposed approach is can be augmented to any CNN.

3.4 Image Perturbations

In the evaluation phase, we experiment with two types of additive noise: Gaussian and uniform. Additive Gaussian noise is part of almost any signal [4], which makes it ideal for mimicking real-life scenarios. An image perturbed with Gaussian noise $\hat{I}_g$ is generated by adding a random value to each pixel (x, y), drawn from a normal distribution $\mathcal{N}$, with a zero mean and a given standard deviation σ to a given image I:

$$\hat{I}_g(x, y) = I(x, y) + \mathcal{N}(0, \sigma) \tag{1}$$

An image perturbed with additive uniform noise $\hat{I}_u$ is created by adding random values drawn from a uniform distribution $\mathcal{U}$, with values between 0 and 1, multiplied by a given weighting value η:

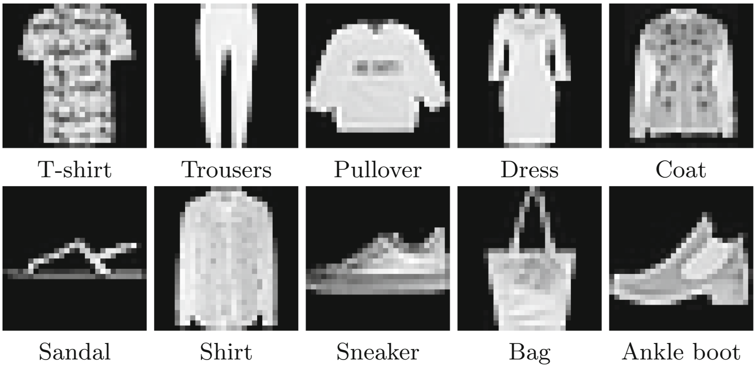
$$\hat{I}_u(x, y) = I(x, y) + \eta \cdot \mathcal{U}(0, 1) \tag{2}$$

4 Experiments and Results

4.1 Data Set

We use the Fashion MNIST data set of Zalando's fashion article images [24]. It has a training and a test set of 60,000 and 10,000 examples, respectively. Each sample is a gray-scale image of 28×28 pixels and belongs to one of the 10 classes as shown in Fig. 7. Since AlexNet accepts images with a size of 227×227 pixels, we resize the images with bi-linear interpolation to these dimensions.

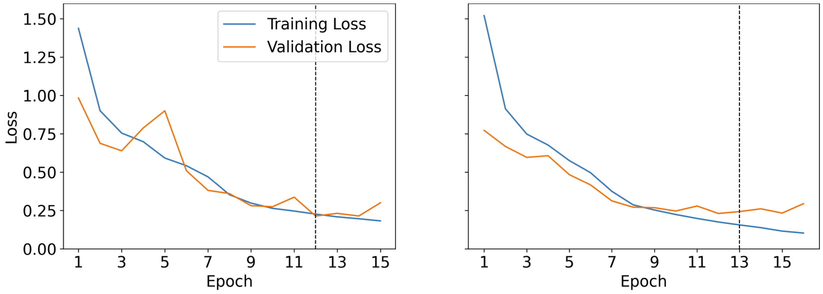
In order to fine-tune the hyper-parameters, we randomly selected a subset of $10,000$ examples in a stratified manner from the training set and used it as a validation set. This resulted in a data set split consisting of $50,000$ images for training, $10,000$ for validation, and $10,000$ for testing.



T-shirt Trousers Pullover Dress Coat

Sandal Shirt Sneaker Bag Ankle boot

Fig. 7. An example of each of the 10 classes in the Fashion MNIST data set.



(a) *CORF-free* (b) *CORF-augmented*

Fig. 8. Training and validation loss without perturbations.

4.2 Experiments

In our experiments, we compare a *CORF-free* pipeline against the proposed *CORF-augmented* pipeline. The hyperparameters of AlexNet used in both pipelines are the same. We used the categorical cross-entropy loss along with the Adam optimizer to train the models, and a batch size of 32. We used an initial learning rate of 0.001 that was decayed at the end of every epoch using an exponential method with a decay rate of 0.96. In order to avoid overfitting, we use a stopping criterium that stops training when the validation accuracy does not improve for three consecutive epochs. This criterium is met at the 12^{th} and 13^{th} epoch for the *CORF-free* and *CORF-augmented* approaches, respectively. From a training point of view, the two pipelines have very similar convergence patterns, depicted in Fig. 8.

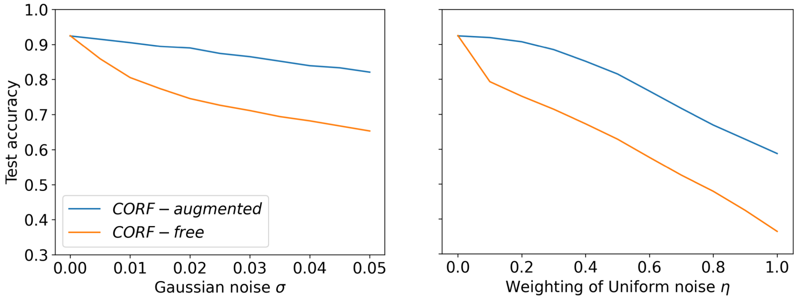
The aforementioned pipelines are implemented as follows: *CORF-free* uses the original Fashion MNIST grayscale images and is used as the baseline. Whereas the *CORF-augmented* pipeline first transforms the given images into CORF contour maps[1] before processing them for classification purposes. Both pipelines are trained with noise-free images and are evaluated with the given test set perturbed by Gaussian and uniform noise of increasing severity.

Figure 9 shows the performance of the *CORF-free* and the *CORF-augmented* models to different levels of Gaussian and uniform noise.

5 Discussion

The results of our experiments confirm our hypothesis on the Fashion MNIST data set, in that an AlexNet trained with CORF contour maps is more robust

[1] The CORF parameters are set as follows. The afferent DoG functions have a standard deviation $\sigma = 5$. As suggested in [1] for $\sigma = 5$, we use two parallel sets of ten center-on and ten center-off collinear DoG functions, whose distances from the center are 34, 18, 9, 5, and 3 pixels. The two parallel sets of center-on and center-off DoG functions are separated by $\beta = 4.0$ pixels and the inhibition factor $\alpha = 5$.



(a) (b)

Fig. 9. Test accuracy for different levels of Gaussian and uniform additive noise.

to unseen additive noise than its counterpart trained with the original grayscale images. As a matter of fact, the proposed *CORF-augmented* approach outperforms that of the conventional *CORF-free* pipeline consistently for different levels of noise. For the maximum severity of noise that we test with, the *CORF-augmented* approach achieves an accuracy of 0.8209 and 0.5874 and reduces the error rate by 48.36 and 35.09 percent for Gaussian and uniform noise, respectively.

Notable is the fact that for noise-free images both pipelines achieve comparable results. The improved robustness, therefore, does not influence the baseline data. To the best of our knowledge, our work is the first to investigate the robustness of CNN-based image classification with CORF push-pull contour maps. In [22] the authors investigated an embedded approach of the push-pull mechanism in CNN models, but it does not involve contour maps as we propose here.

In this preliminary study we investigate only two types of additive noise. Our method, however, has the potential to work on a variety of image corruptions, which we will investigate in future. The conducted experiments only use AlexNet and the Fashion MNIST data set. In future, this work may be extended by investigating other CNNs, such as ZFNet [25], Inception [23] or ResNet [10], other data sets, as well as other types of noise and adversarial attacks. Furthermore, we speculate that using a multi-channel approach to train CNNs can further improve the robustness. The channels may, for instance, include the original color channels of a given image along with CORF response maps with different inhibition strengths.

6 Conclusion

In this work, we show that the proposed pipeline that uses the CORF delineation operator with push-pull inhibition is a promising approach to increase the generalization ability of CNNs. Our experiments included an AlexNet architecture and the Fashion MNIST data set. The proposed *CORF-augmented* pipeline exhibits substantially higher generalization ability for additive Gaussian and uniform noise than a conventional AlexNet without the CORF transformation step.

References

1. Azzopardi, G., Petkov, N.: A CORF computational model of a simple cell that relies on LGN input outperforms the Gabor function model. Biol. Cybern. **106**(3), 177–189 (2012)
2. Azzopardi, G., Petkov, N.: Trainable cosfire filters for keypoint detection and pattern recognition. IEEE Trans. Patt. Anal. Mach. Intell. **35**(2), 490–503 (2013). https://doi.org/10.1109/TPAMI.2012.106
3. Azzopardi, G., Rodríguez-Sánchez, A., Piater, J., Petkov, N.: A push-pull CORF model of a simple cell with antiphase inhibition improves SNR and contour detection. PLoS One **9**(7), e98424 (2014)
4. Boncelet, C.: Image noise models. In: The Essential Guide to Image Processing, pp. 143–167. Elsevier (2009)
5. Da Costa, G.B.P., Contato, W.A., Nazare, T.S., Neto, J.E., Ponti, M.: An empirical study on the effects of different types of noise in image classification tasks. arXiv preprint arXiv:1609.02781 (2016)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
7. Dodge, S., Karam, L.: Understanding how image quality affects deep neural networks. In: 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1–6. IEEE (2016)
8. Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572 (2014)
9. Guo, J., Shi, C., Azzopardi, G., Petkov, N.: Recognition of architectural and electrical symbols by COSFIRE filters with inhibition. In: Azzopardi, G., Petkov, N. (eds.) CAIP 2015. LNCS, vol. 9257, pp. 348–358. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23117-4_30
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
11. Hubel, D.H., Wiesel, T.N.: 8. Receptive fields of single neurones in the cat's striate cortex. In: Brain Physiology and Psychology, pp. 129–150. University of California Press (2020)
12. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning, pp. 448–456. PMLR (2015)
13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. **25**, 1097–1105 (2012)
14. Melotti, D., Heimbach, K., Rodríguez-Sánchez, A., Strisciuglio, N., Azzopardi, G.: A robust contour detection operator with combined push-pull inhibition and surround suppression. Inf. Sci. **524**, 229–240 (2020)
15. Nazaré, Tiago, S., Da Costa, G.B.P., Contato, W.A.., Ponti, M.: Deep convolutional neural networks and noisy images. In: Mendoza, M., Velastín, S. (eds.) CIARP 2017. LNCS, vol. 10657, pp. 416–424. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75193-1_50
16. Neocleous, A., Azzopardi, G., Schizas, C.N., Petkov, N.: Filter-based approach for ornamentation detection and recognition in singing folk music. In: Azzopardi, G., Petkov, N. (eds.) CAIP 2015. LNCS, vol. 9256, pp. 558–569. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23192-1_47

17. Ng, A.Y.: Feature selection, l 1 vs. l 2 regularization, and rotational invariance. In: Proceedings of the Twenty-First International Conference on Machine Learning, p. 78 (2004)
18. Shamsabadi, A.S., Sanchez-Matilla, R., Cavallaro, A.: Colorfool: Semantic adversarial colorization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
20. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
21. Strisciuglio, N., Azzopardi, G., Petkov, N.: Robust inhibition-augmented operator for delineation of curvilinear structures. IEEE Trans. Image Process. **28**(12), 5852–5866 (2019). https://doi.org/10.1109/TIP.2019.2922096
22. Strisciuglio, N., Lopez-Antequera, M., Petkov, N.: Enhanced robustness of convolutional networks with a push-pull inhibition layer. Neural Comput. Appl, pp. 1–15 (2020)
23. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
24. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747 (2017)
25. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
26. Zhai, J., Shen, W., Singh, I., Wanyama, T., Gao, Z.: A review of the evolution of deep learning architectures and comparison of their performances for histopathologic cancer detection. Proc. Manuf. **46**, 683–689 (2020)