Impact of Image Blur on Classification and Augmentation of Deep Convolutional Networks

Matěj Lébl(✉), Filip Šroubek, and Jan Flusser

Department of Image Processing, Czech Academy of Sciences,
Institute of Information Theory and Automation, Prague, Czech Republic
{lebl,sroubekf,flusser}@utia.cas.cz

Abstract. Blur is a common phenomenon in image acquisition that negatively influences the recognition rate of most classifiers. This paper studies the influence of image blurring of various types and sizes on the recognition rate achieved by a deep convolutional network. We confirm that the blur significantly decreases the performance if the network has been trained on clear images only. When the training set is augmented with blurred samples, the recognition rate becomes sufficiently high even if the blur in query images is of different size than the blur used for training. However, this is mostly not true if query images contain blur of a different type from the one used for training.

Keywords: Image recognition · Blur · Augmentation of the training set · Convolutional neural network

1 Introduction

Real digital images are often just degraded versions of the original scene. Image analysis algorithms either must be able to suppress or remove the degradations, or they should be sufficiently robust with respect to them. One of the most common degradations is *blur*, which usually performs smoothing and suppression of high-frequency details of an image. The blur may appear in the image for several reasons caused by different physical phenomena. Motion blur is due to a mutual movement of the camera and the scene (in consumer photography this happens often because of a camera shake), out-of-focus blur appears due to incorrect focus setting, medium turbulence blur is caused by light dispersion and random fluctuations of the refractive index during the acquisition, and sensor blur is a result of a finite size of the detector, to name a few most common blurs. Blur is sometimes introduced intentionally by the photographers to smooth the background, soften the picture and/or to prevent Moire artifacts, but mostly performs an unwanted degradation factor which decreases the image quality and complicates object recognition.

This work was supported by the Czech Science Foundation (Grant No. GA21-03921S) and by the *Praemium Academiae*.

R. Gade et al. (Eds.): SCIA 2023, LNCS 13886, pp. 108–117, 2023.
https://doi.org/10.1007/978-3-031-31438-4_8

Capturing the ideal scene f by an imaging device with the point-spread function (PSF) h and additive random noise n, the observed image g is approximately modeled as a convolution

$$g = f * h + n\,. \tag{1}$$

This linear image formation model, albeit very simple, is a reasonably accurate approximation of many imaging devices and acquisition scenarios.

When recognizing objects in blurred images, we have to take the blur into account and handle it properly. There exist basically four approaches to this problem – image restoration, blur invariants, invariant distance measures, and a learning-based procedure with an augmented database (see Fig. 1).

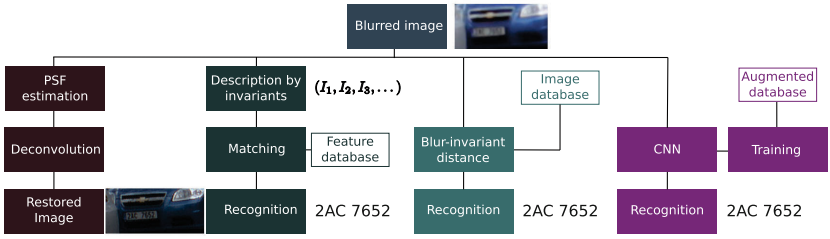

Fig. 1. Four approaches to analysis of blurred images. Image restoration via deconvolution (first branch), description and recognition by blur invariants (second branch), matching by minimum blur-invariant distance (third branch), and a learning-based method (last branch). The restoration yields the recovered image, the other three methods provide the classification result directly without deblurring.

The *image restoration* approach inverts (1) to estimate f from its degraded version g, while the PSF may be partially known or unknown. As soon as f has been recovered, the objects can be described by any standard features and recognized by traditional classifiers. However, image restoration is an ill-posed problem and without any additional constraints infinitely many solutions satisfying (1) exist. To choose the correct one, it is necessary to incorporate a prior knowledge into regularization terms and other constraints. If the prior knowledge is not available, the restoration methods frequently converge to solutions which are far from the ground truth even in the noise-free case. The restoration methods are not employed in this paper, we refer to [1,2] for a survey of current restoration techniques.

In many applications, a full restoration of f is not necessary and can be avoided, provided that an appropriate image representation is used. A typical example is a recognition of objects in blurred images, where an object description robust to blur forms the input to the classifier. This led to introducing the idea of *blur invariants* in 1990s. Roughly speaking, blur invariant I is a functional fulfilling the constraint $I(f) = I(f * h)$ for any h from a certain set of admissible PSFs (see [3] for a survey and further references).

Some authors came up with the concept of *blur-invariant metric*, which allows a blur-robust comparison of two images without constructing the invariants explicitly [4–7].

All the above-mentioned approaches belong to "handcrafted" methods. In accordance with the current trend in using convolutional neural networks (CNN), one may alternatively use deep learning approaches. However, a detailed analysis of robustness of CNN-based classification with respect to blur has not been conducted yet.

It was reported earlier that the blur in query images significantly decreases the CNN performance if the network has been trained on clear images only and that data augmentation (i.e. the extension of the training set with artificially blurred samples) helps to mitigate the performance drop. In this paper, we study the influence of blur of various types and sizes on the recognition rate of CNNs, and analyse different augmentation strategies. The conclusions are that the augmented training set need not to sample the admissible interval of blur sizes densely, however it must contain the largest blur size in the interval. In other words, the generalization property of CNNs with respect to blur is stronger in interpolating rather then extrapolating the training set. Generalization to different types of blur is weak, yet if augmentation with a single blur type is performed, then some blur types generalize better than other. In addition, we compare the CNN-based results with the results achieved by the state-of-the-art "handcrafted" method. To our knowledge, such a systematic study has not been performed yet. All experiments were performed with ResNet18 [8] on reduced datasets of facial images YaleB [9] and general images ImageNet [10].

Both the premise and the conclusion of this paper are intuitive and to be expected. The presence of degradation in the classifying task necessarily affects the result in a negative way and augmenting the training set in a sensible way aids to mitigate the impact of said degradation. The aim of this paper is to describe these effects both qualitatively and quantitatively as well as to offer a guidance on training-set augmentation which yields the best results at minimal extra computational cost.

2 Related Work

Traditional CNNs use pixel-wise representation of images, which is significantly changed if the image has been blurred. So, one may expect a fast drop of performance if the network, trained on clear images only, is fed with blurred test images (our experiments in Sect. 4.1 confirm that). To achieve a reasonable recognition rate, the training set has to be extended using a representative set of degradations of the training images. This process is called *data augmentation*. Common data augmentation methods for image recognition have been designed manually and the best augmentation strategies are dataset-specific [11]. To learn the best augmentation strategy was recently proposed in [12], which searches the space of various augmentation operations and finds a policy yielding the highest validation accuracy. Large-scale data augmentation is, however, time and memory

consuming and for big databases it might be not feasible even on current supercomputers. In our case, the augmentation requires to generate blurred versions of each training image using a proper sampling of the blur space. Since this extends the training set significantly, it is desirable to study the relationship between the number of augmented samples and the recognition rate.

Only few papers have studied the impact of blur on the network recognition performance. Vasiljevic et al. [13] first confirmed experimentally that introduction of even a moderate blur hurts the performance of networks trained on clear images and proposed a fine-tuning of the network to overcome this. A similar study was published by Zhou et al. [14]. Dodge and Karam [15] compared the impact of four types of image degradations – Gaussian blur, contrast deficiency, JPEG artifacts and noise on the CNN-based recognition. They showed that blur is the most significant factor. Most recently, Pei et al. [16] studied the influence of various degradations (local occlusion, lens distortion, aliasing, haze, noise, blur) in detail. Their paper again confirmed that the CNN performance in recognition of blurred images is very low if the training set contains clear images only. However, they did not study the influence of data augmentation with respect to various blur types and sizes and did not advice how it should be performed.

3 Method and Data

The experiments were performed on ResNet18 [8] pre-trained on the ImageNet dataset [10]. The family of ResNet networks score reasonably well in the ImageNet benchmark and the smallest network of size "18" was chosen to speedup the training phase.

For training, we tested two approaches: freezing all except the final layer, and updating the whole net. As retraining the whole model yielded strictly better results, this approach was used in all experiments. The implementation platform was the Python library Pytorch [17].

We considered two classification scenarios. First, the classification classes were 38 faces from the YaleB dataset [9]. Each class was represented by a single image of the face (front view and even lighting) and we ignored other instances in the dataset, such as pose, different lighting, facial expressions, etc. This was done in order to minimize the number of parameters that the network requires to learn and to avoid the effect of other types of degradation on the performance except of the studied blur. Images were resized to 256×256 pixels to fit the ResNet18 default input format. Both the training and test set were augmented by Gaussian white noise and normalised to brightness (to prevent classification based on the intensity of a certain pixel or the whole image). We used this simplified case to analyze and understand the nature of classification of blurred images as well as deduce the optimal augmentation strategy.

In the second scenario, we repeated the same experiments with a subset of the ImageNet dataset to validate the acquired knowledge on a more realistic scenario. We have chosen 5 animal classes - cat, dog, cattle, goat and deer - to have a challenging task. Many of the pictures have fairly similar background or

Fig. 2. Examples of tested blurs: 2× Gaussian, 2× motion, uniform and random blur.

contain couple of different animal species, which makes the classification difficult even without blur (see Fig. 4 for some examples of difficult cases).

Blur augmentation was implemented by a custom function that degrades the input image with blur of varying type (uniform, random, Gaussian, motion) and size ($n \times n, n = 1, \ldots, 256$ px). Examples of the considered blur types and corresponding blurred images are in Figs. 2 and 3.

Fig. 3. Faces used in the experiment. From left to right: no blur, motion, Gaussian, uniform and random blur; from top to bottom: blur size 15, 45 and 65 pixels.

4 Experiments

We have conducted two sets of experiments. The first one verifies that the classification network performs poorly on the blurred test data when the data augmentation is not carried out. The second part focuses on the process of augmentation itself, specifically we explore the sampling strategy of the space of admissible blurs that yields high accuracy.

As mentioned earlier, we prepared two dataset: 38 single-image classes from the YaleB dataset and 5 classes from the ImageNet dataset.

In the first dataset, we have strong similarity between classes, but zero variety inside them. With this setup, we alleviate the "learning class variety" problem and can focus solely on the effect of blur. Thanks to the simplicity of the task, the baseline accuracy is 100 % and we can use substantial blur with fine sampling to smoothly observe the decline in accuracy under various scenarios.

The second dataset is more challenging. Classes are fairly similar to each other (cattle - goat - deer) and have common background (indoors: cat - dog,

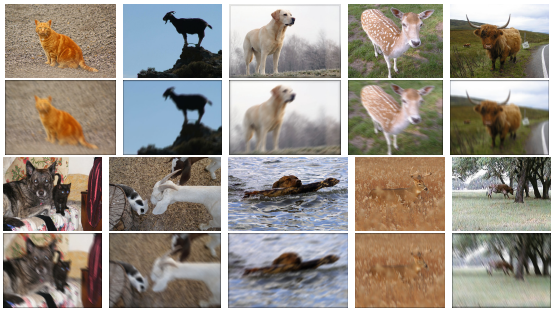

Fig. 4. Animal images and their blurred versions used in the experiment. Examples of one easy (top) and one challenging/ambiguous (bottom) image from each class. Degraded versions were blurred by Gaussian blur of size 15 pixels. By rows: cats, goats, dogs, deer, cattle.

fields/meadows: cattle - deer) with hundreds of different images in each class. The baseline accuracy in this setting is 86 %. Albeit the impact of blur is more sever in this case, we can still identify the same behaviour as in the first, simplified case scenario, yet on a smaller scale.

4.1 Different Types of Blur

Augmenting the training set is necessary for classification of degraded images and we explore various scenarios of training and testing on different types of blur.

In the first experiment, we train the network on a clear dataset and measure how large the blur can be without hindering the accuracy. We use all four blur types in the test set. Figure 5 (blue line) clearly shows that for blurs larger than 7×7 pixels, the accuracy starts to quickly decline.

Similar trend, albeit not so profound, holds for the ImageNet dataset (orange line). Note that the slower decline in accuracy is partially caused by the smaller number of classes - with random assignment of classes, the ImageNet dataset would have $\sim 20\%$ accuracy opposed to $\sim 2.5\%$ for the face dataset.

Next, we are interested in the network performance when trained on one specific type of blur and evaluated on other types of blur. We prepare four sets of training and test data, each degraded by a different type of blur (Gaussian, motion, random and uniform), and train and test all combinations. Results summarized in Table 1 show that the random blur is a good approximation of the uniform one and vice versa, while both of them are poor approximation to the Gaussian and motion blur. Likewise, the Gaussian and motion blur poorly approximate the uniform and random blur. Interesting finding is that while the motion blur approximates the Gaussian blur relatively well, it is not true the other way around. Similar asymmetry holds also for the Gaussian and uniform blur.

The "PM" column of the table contains the results of a handcrafted projection method, originally proposed by Vageeswaran et al. [6] and improved by Lebl et al. in [7], where the images are classified by minimum blur-invariant distance. Please note that albeit the handcrafted projection method shows excellent results, it is not invariant to geometric transformations or any other degradation apart from translation and convolution, e.g. a slightly rotated query image will not be recognized. Invariance to these transformations can be in theory implemented by brute force, which is in a way similar to data augmentation in learning-based methods, yet the most limiting factor is the nonexistence of intraclass generalization, e.g. different types of dogs will not be classified as members of the same class "dog". From this perspective, the projection method is not a rival to learning-based methods if the intraclass variations are significant and not parametrizable.

The results on the diagonal demonstrate that with the right augmentation of the training set, the network performs equally well as the handcrafted method. However, the main difference appears if the blur type is not known a priori. While the network cannot be trained properly, the performance of the handcrafted projection method is not affected.

Repeating the experiment on ImageNet dataset shows in Table 1 similar clustering of uniform-random and Gaussian-motion blurs and verifies that information about blur type is crucial in maximizing the effect of augmentation.

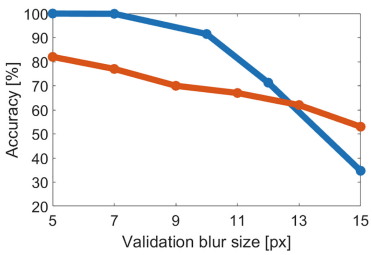

Fig. 5. The recognition rate [%] on images blurred with different blur size. The network was trained on clear (sharp) images only. Blue line is for the simplified YaleB dataset, orange for the ImageNet dataset. (Color figure online)

4.2 Interpolation and Extrapolation

The critical question is the amount of augmentation of the training set necessary for maintaining high accuracy. We evaluate how important is to know the maximum possible size of blur and how dense the sampling of the space of all possible degraded images must be. In other words, we are interested in the generalization property of extrapolation and interpolation in the space of blurs.

Table 1. The recognition rate [%] for the YaleB and ImageNet dataset, only one type of blur (125×125px/19×19px) was used for training (column) and testing was done on 4 sets of images degraded with 4 blur types respectively (row). Reference result (PM) of the handcrafted method [7] are in the middle column.

Train/Test	Yale B				PM	ImageNet			
	uniform	random	Gauss	motion		uniform	random	Gauss	motion
uniform	99,63	99,56	46,62	38,57	97,63	72	69	67	58
random	99,57	99,59	29,18	45,54	100	69	71	69	60
Gauss	24,84	20,36	99,88	55,74	100	67	67	70	70
motion	39,97	34,32	82,04	99,91	100	63	60	68	74

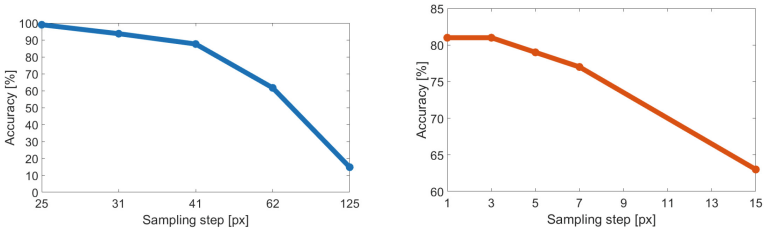

Fig. 6. The recognition rate [%], tested on the YaleB-left (ImageNet-right) dataset randomly degraded by blur up to size 125×125 (15×15). The training set was augmented with different sampling step of blur sizes from the interval $[1, 125]$ ($[1, 15]$). Original images without blur were always included in the training set.

We use all four blur types in the following experiments as no single blur type is good enough to approximate all other blurs. To test the interpolation capability of the network, we use test images degraded by all blur sizes up to 125×125 px while training on data augmented only by blurs of size 1(no blur), $k, 2k, ..., nk \leq 125 \times 125$ px, where k is the sampling step. The total number of training data is the same for all tested sampling steps. This ensures that the change in accuracy is caused by different sampling and not by a larger/smaller training set. From the results in Fig. 6-top, we can see that the network is able to cover gaps in sampling up to 25 px without losing accuracy. This means that we can correctly classify images degraded by blurs that are roughly 12 px apart from the nearest blur used in the augmentation of the training set.

For the ImageNet dataset, we set the maximal blur to 15×15 px and use much finer sampling steps. Even in this down-scaled scenario, we can see in Fig. 6-bottom that it is not necessary to include all admissible blur sizes. Sampling step 3 is sufficient to achieve the same results as when the exhaustive sampling with step 1 is used.

For extrapolation, we train the network on images degraded by blurs up to size $N \times N$ and test on a set of images degraded by a blur of size exactly 75×75 px (YaleB dataset). This simulates a situation when we underestimate the

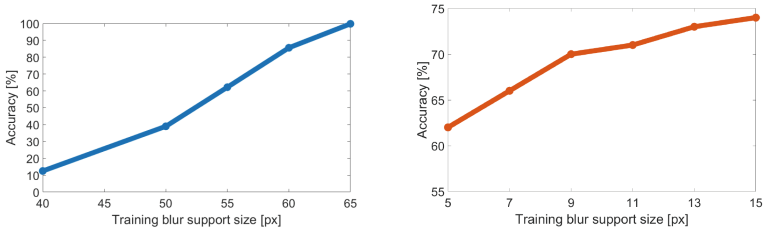

Fig. 7. The recognition rate [%], trained on the YaleB-left (ImageNet-right) dataset augmented by blur of varying sizes and tested on data with blur of size exactly 75 (15) and SNR= 15 dB (SNR= 20 dB).

extent of possible degradation. Results in Fig. 7-top show that the extrapolation capability is limited to blurs of approximately 10 px outside the training set. Augmenting the training set with the largest expected blur is thus important for preserving good accuracy. Note that the results are in agreement with the previous interpolation experiment where the minimum distance necessary for 100% accuracy was 12 px between blur samples.

For the ImageNet dataset, the test set contains images degraded by blur exactly 15×15 px and the training set is augmented with blurs up to size $N \times N$, $N \in \{5, 7, 9, 11, 13, 15\}$ px. The results in Fig. 7-bottom show that the ability to extrapolate is almost lost. Also note the difference between the best achieved accuracy on Figs. 6-bottom and 7-bottom. This is caused by the difference in test sets: in the former, we use blurs of size 1 (no blur) to 15 (maximal blur) in the latter we use exclusively blurs of size 15.

5 Conclusion

We confirmed that even for a very simple task, augmenting the training set with blur is necessary, when one wants to recognize blurred query images by CNNs.

For the augmentation itself we advice either to use the same blur type as is expected in query images (if this information is available) or to include as many different blur types as possible because the augmentation by a single blur type is insufficient if it is different from the blur in queries.

As for choosing the right blur sizes for augmentation, it is sufficient to sample uniformly with reasonably small (≤ 25 px) step up to the largest expected blur. Underestimating the size of possible blur is tolerated only up to small margin - roughly half of the recommended sampling step.

We validated that the trend as well as the training strategy holds for more realistic scenario although (as one would expect) we have to scale down the blur sizes accordingly.

Since the augmentation of the training set is a costly process, looking for efficient alternatives is a challenge for the near future. One of the choices we plan to study is to incorporate handcrafted blur-invariant features [18] into the CNNs. This could reduce or even remove the necessity of data augmentation by blurred images.

References

1. Campisi, P., Egiazarian, K.: Blind Image Deconvolution: Theory and Applications. CRC, Boca Raton (2007)
2. Rajagopalan, A.N., Chellappa, R.: Motion Deblurring: Algorithms and Systems. Cambridge University Press, Cambridge (2014)
3. Flusser, J., Suk, T., Zitová, B.: 2D and 3D Image Analysis by Moments. Wiley, Chichester (2016)
4. Zhang, Z., Klassen, E., Srivastava, A.: Gaussian blurring-invariant comparison of signals and images. IEEE Trans. Image Process. **22**(8), 3145–3157 (2013)
5. Gopalan, R., Turaga, P., Chellappa, R.: A blur-robust descriptor with applications to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **34**(6), 1220–1226 (2012)
6. Vageeswaran, P., Mitra, K., Chellappa, R.: Blur and illumination robust face recognition via set-theoretic characterization. IEEE Trans. Image Process. **22**(4), 1362–1372 (2013)
7. Lébl, M., Šroubek, F., Kautský, J., Flusser, J.: Blur invariant template matching using projection onto convex sets. In: Proceedings of 18th International Conference on Computer Analysis of Images and Patterns, CAIP 2019, pp. 351–362 (2019)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, June 2016, pp. 770–778. IEEE Computer Society (2016)
9. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intelligence **23**(6), 643–660 (2001)
10. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009)
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (2017)
12. Cubuk, E.D., Zoph, B., Mané, D., Vasudevan, V., Le, Q.V.: Autoaugment: learning augmentation strategies from data. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 113–123 (2019)
13. Vasiljevic, I., Chakrabarti, A., Shakhnarovich, G.: Examining the impact of blur on recognition by convolutional networks. arXiv preprint arXiv:1611.05760 (2016)
14. Zhou, Y., Song, S., Cheung, N.: On classification of distorted images with deep convolutional neural networks. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1213–1217. IEEE (2017)
15. 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)
16. Pei, Y., Huang, Y., Zou, Q., Zhang, X., Wang, S.: Effects of image degradation and degradation removal to CNN-based image classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(4), 1239–1253 (2021)
17. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 32, pp. 8024–8035. Curran Associates Inc. (2019)
18. Flusser, J., Suk, T., Boldyš, J., Zitová, B.: Projection operators and moment invariants to image blurring. IEEE Trans. Pattern Anal. Mach. Intell. **37**(4), 786–802 (2015)