sciendo

Int. J. Appl. Math. Comput. Sci., 2018, Vol. 28, No. 4, 735–744
DOI: 10.2478/amcs-2018-0056

amcs

IMPACT OF LOW RESOLUTION ON IMAGE RECOGNITION WITH DEEP NEURAL NETWORKS: AN EXPERIMENTAL STUDY

MICHAŁ KOZIARSKI [a,*], BOGUSŁAW CYGANEK [a]

[a]Department of Electronics
AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Kraków, Poland
e-mail: {michal.koziarski,cyganek}@agh.edu.pl

Due to the advances made in recent years, methods based on deep neural networks have been able to achieve a state-of-the-art performance in various computer vision problems. In some tasks, such as image recognition, neural-based approaches have even been able to surpass human performance. However, the benchmarks on which neural networks achieve these impressive results usually consist of fairly high quality data. On the other hand, in practical applications we are often faced with images of low quality, affected by factors such as low resolution, presence of noise or a small dynamic range. It is unclear how resilient deep neural networks are to the presence of such factors. In this paper we experimentally evaluate the impact of low resolution on the classification accuracy of several notable neural architectures of recent years. Furthermore, we examine the possibility of improving neural networks' performance in the task of low resolution image recognition by applying super-resolution prior to classification. The results of our experiments indicate that contemporary neural architectures remain significantly affected by low image resolution. By applying super-resolution prior to classification we were able to alleviate this issue to a large extent as long as the resolution of the images did not decrease too severely. However, in the case of very low resolution images the classification accuracy remained considerably affected.

Keywords: image recognition, deep neural networks, convolutional neural networks, low resolution, super-resolution.

1. Introduction

Deep neural networks, and more specifically convolutional neural networks, emerged in recent years as a methodology of choice in various computer vision problems. They have been successfully used in numerous tasks, such as semantic segmentation (Long *et al.*, 2015), facial point detection (Sun *et al.*, 2013), dense captioning (Johnson *et al.*, 2016), biomedical image segmentation (Ronneberger *et al.*, 2015) and image restoration (Mao *et al.*, 2016), to name just a few. Perhaps most importantly, deep neural networks have been used in the image recognition task, achieving superhuman visual pattern recognition in many controlled competitions (Schmidhuber, 2015).

Such competitions, especially the ImageNet challenge (Russakovsky *et al.*, 2015), were the driving force behind the recent progress in neural-based models. Numerous novel architectures of deep neural networks (Krizhevsky *et al.*, 2012; Simonyan and Zisserman, 2014; He *et al.*, 2016) have gained their popularity based on the results achieved on the ImageNet benchmark. However, ImageNet consists of fairly high quality images, while in practical applications we are often faced with low quality of images and the presence of factors such as low resolution, noise, blur, compression artifacts and a low dynamic range. While human examiners are, to a large extent, resilient to such factors during image recognition, it is unclear to what degree low image quality affects the performance of deep neural networks.

In this paper we evaluate the impact of low resolution on the image recognition task, considering several most notable neural architectures of recent years, that is, AlexNet (Krizhevsky *et al.*, 2012), VGGNet (Simonyan and Zisserman, 2014) and ResNet (He *et al.*, 2016). Furthermore, we examine application of the neural-based super-resolution method, VDSR (Kim *et al.*, 2016), to improve classification accuracy for low resolution images.

*Corresponding author

2. Related work

Poor image quality has been recognized as an important aspect influencing the performance of deep neural networks in computer vision tasks. Various factors influencing image quality have been considered in the context of classification accuracy of deep neural networks. Dodge and Karam (2016) evaluated the impact of blur, noise, contrast and JPEG compression on the performance of several neural architectures. Karahan *et al*. (2016) examined blur, noise, compression artifacts, color distortions and occlusion in the context of the face recognition task. Vasiljevic *et al*. (2016) performed a more thorough analysis of different types of blur, at the same time considering different methods of dealing with this type of distortion. Koziarski and Cyganek (2017) evaluated the impact of the presence of different types of noise and the effectiveness of different approaches to cope with these phenomena. Sanchez *et al*. (2016) examined the influence of illumination quality and contrast measures. Several papers evaluating the impact of low image quality on different types of classification algorithms can also be found (da Costa *et al*., 2016; Dutta *et al*., 2012). In particular, Zou and Yuen (2012) considered the problem of a very low resolution. However, to the best of our knowledge, no research on the impact of low image resolution on classification accuracy of deep neural networks has been conducted yet.

Applying super-resolution prior to image recognition is an intuitive approach to mitigating the negative impact of low image resolution on classification accuracy. The problem of image super-resolution has a rich research history in the computer vision community. Currently, the most dominant paradigm relies on training the model based on the available data, the methodology that can be traced back to the work of Freeman *et al*. (2002). This family of methods can be based on various machine learning algorithms, such as nearest neighbor approaches (Freeman *et al*., 2002), manifold learning (Chang *et al*., 2004), dictionary learning (Yang *et al*., 2010), locally linear regression (Timofte *et al*., 2014) and random forests (Schulter *et al*., 2015), to name just a few. However, with the advent of deep learning, convolutional neural networks began to outperform methods based on other machine learning techniques.

The SRCNN (Dong *et al*., 2014) was the first of convolutional-based neural architectures achieving the state-of-the-art results in the domain, afterwards improved by Dong *et al*. (2016). Since then, numerous variants of convolutional neural networks have been proposed for the image super-resolution task. Out of them, the most notable are the following: a sub-pixel convolutional network proposed by Shi *et al*. (2016), designed as a real-time method, a generative adversarial network proposed by Ledig *et al*. (2016), as well as a fully convolutional architecture taking advantage of residual connections, the VDSR network proposed by Kim *et al*. (2016), which was later extended by Tai *et al*. (2017) to enable recursive connections. The notion of using residual connections was also further examined by Lim *et al*. (2017), who proposed novel single-scale and multi-scale architectures. However, despite the abundance of super-resolution research, to the best of our knowledge, the possibility of using a super-resolution technique as a form of low resolution image preprocessing in the task of classification with deep neural networks has not been examined before. Thus, we fill this gap with the methods described in this paper.

Finally, applying the super-resolution is not the only method of dealing with low resolution in the context of image recognition. Another notable approach was described by Peng *et al*. (2016), who proposed transferring the fine-grained knowledge obtained from high resolution data to the task of low resolution image classification.

3. Background

In this section we provide a brief overview of deep neural networks in the context of image recognition and super-resolution tasks. We provide the problem definitions, describe the methodology of using neural networks and highlight the most relevant neural architectures. A more detailed description of deep neural networks can be found in the work of Goodfellow *et al*. (2016).

3.1. Image recognition with deep neural networks. Let us denote the input image by X and the associated label, belonging to a set of a fixed length, by y. The goal of the image recognition task is providing a mapping $f(X)$ best approximating y for a given X. By far the most prevalent approach to the image recognition problem is the data-driven paradigm: instead of explicitly defining the mapping $f(X)$, we propose a parametrized family of models $f(X|\theta)$, with θ being a set of trainable parameters. Based on the available data, we later optimize the values of θ. Deep neural networks are a particular type of a trainable model. They consist of building blocks called neurons, which are basic computational units of neural networks. The neurons are grouped into consecutive layers, which form a complete model. Based on the arrangement of neurons within layers and layers within the whole network, we define the function space used in the optimization procedure. The choice of the arrangement of the neural network is therefore essential to achieve good performance in the image recognition task, and a significant amount of research has been dedicated to evaluating various architectures. Below we describe three of the most notable models of neural networks, used later in the experimental analysis.

AlexNet (Krizhevsky *et al*., 2012) was an architecture developed by Alex Krizhevsky and co-workers, as well as the winner of the 2012 edition of the ImageNet ILSVRC challenge (Russakovsky *et al*., 2015). It was a neural architecture that popularized the usage of convolutional neural networks in the image recognition task, achieving considerably better results than the state-of-the-art methods of the time. It consists of 60 million parameters, grouped into five convolutional layers, some of which followed by a pooling layer, and three fully-connected layers. At the time, AlexNet was one of the largest neural networks applied. It was also one of the first networks using rectified linear units as nonlinearity.

VGGNet (Simonyan and Zisserman, 2014) was a neural architecture introduced by the Visual Geometry Group of the University of Oxford, and a contender in the 2014 edition of the ILSVRC challenge. Compared to AlexNet, VGGNet consists of a significantly higher number of parameters: 144 million for the largest possible variant. The main contribution of VGGNet was showing the importance of the depth of convolutional neural networks. In total, the largest version of VGGNet consisted of 19 weight layers. Filters with a very small receptive field, with the size of 3×3, were used in the convolutional layers to decrease the number of parameters without limiting the depth. To make the learning feasible, the training was conducted in stages: instead of optimizing a large, randomly initialized network, shallower versions were trained first, and afterwards used to initialize the weights of deeper networks.

ResNet (He *et al*., 2016) was a neural network that won the 2015 edition of the ILSVRC challenge. It introduced the concept of residual connections: instead of learning unreferenced functions, in the residual framework the functions with reference to the layer inputs are learned. This allowed the authors to successfully train a networks with up to 152 layers. Furthermore, since only a single fully connected layer was used, the overall number of parameters was significantly reduced to only 60 million.

3.2. Image super-resolution. Let us denote by X_H a ground truth, high resolution image, and by X_L its low resolution counterpart. The goal of the super-resolution task is, given X_L, estimating $g(X_L)$ resembling the ground truth image as closely as possible. The problem is inherently ill-posed, since multiple solutions may exist for a given input image. This issue is particularly pronounced for very low resolution images, for which the space of possible source, high resolution images is larger.

In the context of the image recognition task, super-resolution can be viewed as a preprocessing technique for low resolution images. In general, we are concerned with the task of low resolution image recognition, that is, the optimization of $f(X_L|\theta)$ to find the best approximation of y. However, in the practical setting we may be faced with several limitations that may make the direct optimization of $f(X_L|\theta)$ impossible. First of all, training deep neural networks is usually a computationally expensive procedure. In real applications we may not be able to afford the cost of training a complete model. Pre-trained models can be used in such cases, but they are usually trained on relatively high quality data and may not be suitable for the recognition of low resolution images. Secondly, deep neural networks used for image recognition usually require large amounts of data. We may not know the exact resolution of the images of interest, which would make artificially lowering the image resolution impossible, and may not have enough real data to train the model without that operation. Using a super-resolution technique prior to classification to approximate X_H using $g(X_L)$ and afterwards approximating y using $f(g(X_L)|\theta)$, with θ trained based on X_H, can be a suitable approach in such cases.

The method we chose for the super-resolution task is VDSR, a neural architecture proposed by Kim *et al*. (2016). It is a fully-convolutional model with a conceptually simple structure: it consists of 20 convolutional layers, each containing 64 filters of size 3×3. Furthermore, VDSR takes advantage of residual learning, introducing a skip connection between the input and the output of the network. As can be seen, these are the exact main attributes of the most successful image recognition networks: significant depth, a small filter size, relying on the convolutional layers and the presence of skip connections. In addition, the VDSR network introduced the concept of multi-scale learning, that is training a single model on data with varying levels of low resolution. The VDSR network displays great performance in the super-resolution task, at the same time being simple and fast to train.

4. Experimental study

To assess the impact of low resolution on the task of image recognition with deep neural networks, an empirical study was performed. Firstly, we measured classification accuracy of various neural architectures on images with a varying level of low resolution. Secondly, we evaluated the performance of a state-of-the-art super-resolution method, VDSR, to establish its suitability as a form of preprocessing in the image recognition task. Finally, we integrated super-resolution into the image recognition pipeline in an attempt to improve classification accuracy. In the remainder of this section we give a detailed description of the conducted experimental procedure, present the achieved results and state our conclusions.

4.1. Impact of low resolution on the image recognition task. We began our evaluation by measuring how the quality of images affects the classification accuracy in the situation in which low resolution is not accounted for. To this end, we examined three of the most significant neural architectures of recent years: AlexNet (Krizhevsky *et al.*, 2012), 16-layer VGGNet (Simonyan and Zisserman, 2014) and 50-layer ResNet (He *et al.*, 2016). Our goal was, first of all, to evaluate how significant is the decrease in the accuracy due to low image resolution. Furthermore, we aimed to establish whether the observed trends are similar across the different neural architectures, or if some of the models are more resilient to poor image quality.

Given an image and a scale factor (SF), we artificially decreased the resolution of the luminance channel of the image. We manipulated only the luminance channel to ensure consistency with the super-resolution framework described by Huang *et al.* (2015). Firstly, we reduced the size of the luminance channel to $1/SF$ of the original, and afterwards we increased it back using bicubic interpolation. An example of an image with reduced resolution is shown in Fig. 2.

We conducted our evaluation on the ImageNet (Russakovsky *et al.*, 2015) dataset. Specifically, we used a subset of images provided during the Large Scale Visual Recognition Challenge 2012 (ILSVRC2012). ImageNet is a standard, publicly available benchmark, commonly used to evaluate the performance of convolutional neural networks in the image recognition task. It consists of 1.2 million training and 50 thousand validation images, grouped into 1000 categories. Throughout the performed evaluation, suggested images were used during the training of the models, whereas the reported results were based on the performance on the validation data. Instead of training the models from scratch, the weights provided by the authors of the corresponding papers were used in every case. The experiment itself was implemented in the Python programming language, using the TensorFlow (Abadi *et al.*, 2016) machine learning library. The resulting code was made available at `https://github.com/michalkoziarski/LowResCNN`.

During the examination we measured the impact of the SF in $\{1, 2, \ldots, 8\}$ on both top-1 and top-5 classification accuracy, with top-k accuracy being defined as a fraction of the images for which the ground truth image label was included in the k most probable predictions of the classifier. We present the results of this part of the experimental evaluation in Fig. 2. As can be seen, low resolution of the images significantly decreases both top-1 and top-5 classification accuracy of the neural networks considered. Even in the case of the mildest image quality deterioration, represented by an SF equal to 2, the drop in accuracy was noticeable. In the case of top-1 accuracy it ranged from 3.88 percentage points for AlexNet, through 2.31 percentage points for VGGNet, up to 2.28 percentage points for ResNet. In the case of top-5 accuracy the observed drop was 3.15, 1.53 and 1.69 percentage points for AlexNet, VGGNet and ResNet, respectively. The observed drop in performance was much more significant at higher deterioration levels. For the SF equal to 8, top-1 accuracy decreased by 30.32, 30.64 and 27.12 percentage points, whereas top-5 accuracy by 33.63, 25.89 and 20.80 percentage points, respectively, for AlexNet, VGGNet and ResNet.

As can be seen, the observed trends are similar between the discussed architectures of neural networks. The relation between classification accuracy and the SF of low resolution images is close to being linear, in all the cases. The disproportion between top-1 and top-5 accuracy remains relatively stable for different values of the SF. For low values of the SF, both VGGNet and ResNet display a similar drop in performance. However, for a higher SF, ResNet was the most resilient model, especially when top-5 classification accuracy was considered. In general, even the state-of-the-art architectures of neural networks remain significantly affected by the issue of low image resolution. Visually examining images with lowered resolution, presented in Fig. 1, even small values of the SF that should not affect human recognition capabilities noticeably affect classification accuracy of neural networks.

4.2. Image super-resolution with the VDSR network. The goal of the latter part of the experimental study was evaluating the possibility of applying image super-resolution prior to recognition, with the hope of improving classification accuracy for low resolution images. As alluded to previously, the VDSR (Kim *et al.*, 2016) neural architecture as a super-resolution method was chosen, since it displayed a state-of-the-art performance in the SR task. Furthermore, since both the super-resolution model as well as the classification model are neural networks, it presents the practical benefit of the possibility of treating the whole system as a single, large network. This, in turn, allows the usage of operations such as fine-tuning the combined network. In the remainder of this section we experimentally evaluate the super-resolution capabilities of the model considered, that is, the VDSR neural network with an adjusted training procedure. We describe the adjustments made, experimental set-up and the achieved results. The implementation of the VDSR model used in this part of the experimental study was made publicly available at `https://github.com/michalkoziarski/VDSR`.

Similarly to Kim *et al.* (2016), for the training of the model we used a combination of 91 images provided by Yang *et al.* (2010) and 200 images from the Berkeley

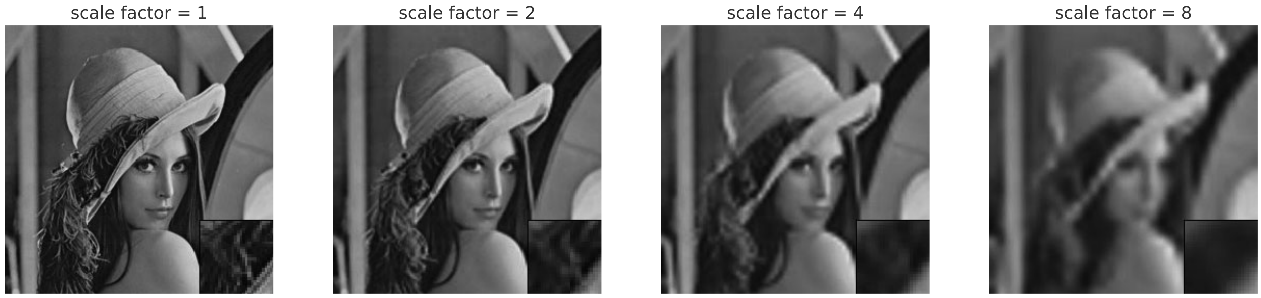

Fig. 1. Example of an image with the resolution artificially lowered using bicubic interpolation (original image on the far left).

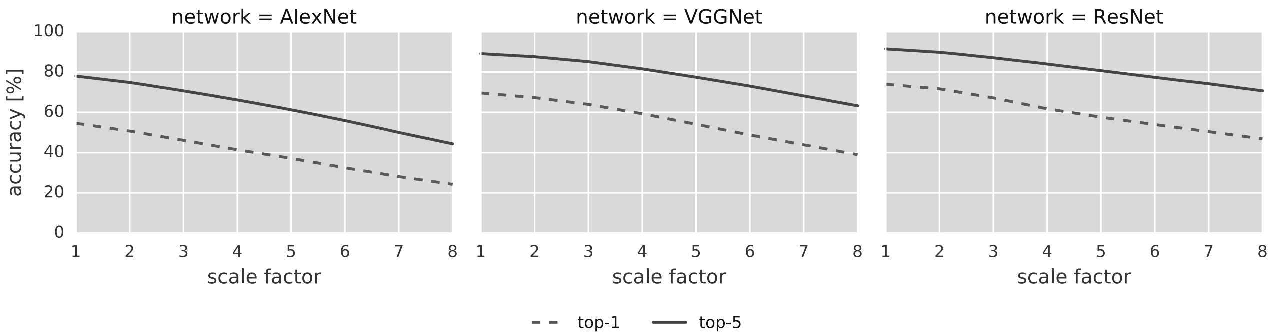

Fig. 2. Impact of low image resolution on classification accuracy with various convolutional neural networks.

Segmentation Dataset (Martin *et al*., 2001). The data were further augmented applying rotation and flipping, which effectively increased the size of the dataset 8 times. For the evaluation of the models' performance we used four different benchmarks: Set5 (Bevilacqua *et al*., 2012), Set14 (Zeyde *et al*., 2010), 100 test images from the Berkeley Segmentation Dataset (B100) (Martin *et al*., 2001) and Urban100 (Huang *et al*., 2015). When combined, the benchmarks considered consist of diverse images from a broad distribution, including images of people, animals, buildings and inanimate objects.

Compared with Kim *et al*. (2016), we adjusted the proposed training procedure. Most importantly, we discovered that the training is susceptible to the initialization of the weights, and the final performance can vary depending on the starting point of the optimization. We were able to stabilize the learning process to a large extent by using the Adam optimizer (Kingma and Ba, 2014) instead of the stochastic gradient descent, which allowed easier reproducibility of the results. We used the values of $\beta_1 = 0.9$ and $\beta_2 = 0.999$ for the Adam algorithm. Furthermore, we were unable to avoid the issue of exploding gradients when starting the training with very high learning rates. This was the case even though we were using the gradient clipping procedure described in the original paper. Still, we clipped the gradients to the range of $[-\theta/\gamma, \theta/\gamma]$, with γ being the current learning rate, and θ the parameter of the gradient clipping procedure. In our experiments we set $\theta = 0.5$, while in the original paper that choice was not mentioned. Because of the described issues with exploding gradients, we began the optimization with a learning rate equal to 0.0001. We further decreased it by a factor of 10 every 20 epochs. In total, we trained the model for 60 epochs. We did not adjust the remaining parameters of the training procedure, that is, batch size of 64, patch size of 41 and weight decay of 0.0001. The architecture of the network remained unchanged as well. The training and evaluation times of the model were comparable to those described in the original paper.

We present the average values of the peak signal-to-noise ratio (PSNR) for different benchmarks in Table 1. We duplicated the results achieved by the reference methods from Kim *et al*. (2016) while adjusting the values for the work of VDSR based on our experiments. Despite the adjustments made, which improved the performance of the algorithm, we were still unable to achieve the exact results reported in the original paper, overall observing slightly worse performance. The drop in performance was, however, negligible, and VDSR still outperformed the reference methods in every single case. Since the code necessary to train the model was not made available together with the original paper, for the sake of the reproducibility we decided to use our model during further evaluation. To assess the statistical significance of the observed results we also conducted a Wilcoxon signed-rank test. The results observed for VDSR were statistically significantly different than those observed for all of the reference methods at a significance level of 0.005.

A sample image from Set5 after applying super-resolution is presented in Fig. 3. As can be seen, applying the VDSR algorithm produces sharp edges and eliminates the blurring associated with low resolution images. However, especially for larger values of the SF, loss of some of the detail remains unavoidable. VDSR, while producing visually pleasing images, is unable to restore the full information that could later be used during classification.

4.3. Applying super-resolution prior to classification. In the final stage of the conducted experimental study we examined whether applying super-resolution prior to classification can improve classification accuracy for low resolution images. We considered all of the previously used neural network architectures, that is, AlexNet, VGGNet and ResNet, and the SF in $\{1, 2, 3, 4\}$. We used the VDSR neural network, trained according to the procedure described in the previous section, as the super-resolution algorithm.

Visualization of the trends observed during this part of the experimental study is presented in Fig. 4, whereas the complete numerical results are presented in Table 2. As can be seen, applying super-resolution as a form of image preprocessing led to an improvement in the performance in every single case. For the SF equal to 2 we were able to achieve performance close to the one observed on the original, undistorted data. When top-1 accuracy was considered, compared with the case of undistorted data, the observed drop in classification accuracy was equal to 0.47 percentage points for AlexNet, 0.45 for VGGNet and 0.43 for ResNet. This means that applying super-resolution eliminated 87.92%, 80.59% and 81.15% of the performance drop, i.e., the difference between the baseline and the low-resolution affected data, for AlexNet, VGGNet and ResNet, respectively. Similar trends were observed for top-5 accuracy: for AlexNet, the drop in performance was reduced to 0.35 percentage points, meaning that 88.76% of the accuracy drop caused by low image resolution was eliminated. For VGGNet the observed values were 0.25 percentage points and 83.46%, whereas for the ResNet 0.35 percentage points and 79.14%.

In the case of a severely reduced image resolution, that is when the SF was equal to 4, the degree to which the classification accuracy was restored due to applying the super-resolution was significantly lower. For the top-1 accuracy, 29.17% of the performance drop caused by the low resolution was eliminated for the AlexNet, 31.36% for the VGGNet and 35.84% for the ResNet. Still, compared with the case in which the low image resolution was not accounted for, applying super-resolution increased the accuracy by 3.84 percentage points for the AlexNet, by 3.24 for VGGNet and by 4.37 for the ResNet. Once again, the observed trends were similar for the top-5 accuracy. In that case, applying the super-resolution eliminated 30.52% of the performance drop for AlexNet, 37.30% for VGGNet and 32.59% for ResNet. To assess the statistical significance of the observed results, a Wilcoxon signed-rank test was conducted. The null hypothesis that the classification accuracy observed after applying super-resolution prior to classification and the results observed for low resolution images come from the same distribution was rejected at the 0.001 significance level.

In summary, applying super-resolution as a form of image preprocessing is a suitable approach for low resolution images as long as the decrease in their quality is not too significant. For the case of an SF equal to 2, by applying super-resolution we were able to obtain classification accuracy close to that observed on the undistorted data. This was not the case for higher levels of distortion. However, while we were unable to restore classification accuracy close to that observed on the original images, a noticeable improvement was still achieved when compared with the case in which low resolution was not accounted for in any way.

5. Conclusions

In this paper we experimentally evaluated the impact of low resolution on the task of image recognition with deep neural networks. We measured the effect of artificially induced low resolution on classification accuracy of the most notable neural architectures of recent years. Furthermore, we evaluated the possibility of applying super-resolution as a form of preprocessing to increase classification accuracy for low resolution images. The main findings of this paper are the following:

- Low image resolution, when not accounted for, may significantly decrease classification accuracy of deep neural networks. This is true even if the decrease in image resolution is relatively mild: a noticeable drop in performance was still observed for the lowest analysed level of low resolution, arguably difficult to spot with a naked eye in many cases. This stands in a stark contrast with the supposed "superhuman" (Schmidhuber, 2015) capabilities of deep neural networks in the image recognition task.

- The observed trends were similar across different architectures of neural networks. Based on our observations, the models achieving higher accuracy

Table 1. Average PSNR values for various super-resolution methods. Results replicated from the work of Kim *et al.* (2016), with the updated values for the VDSR method, based on our own experiments. Best performance in bold.

Dataset	Scale	Bicubic	A+ (Timofle *et al., 2014)*	RFL (Schulter *et al., 2015)*	SelfEx (Huang *et al., 2015)*	SRCNN (Dong *et al., 2014)*	VDSR (Kim *et al., 2016)*
Set5	$\times 2$	33.66	36.54	36.54	36.49	36.66	**36.79**
	$\times 3$	30.39	32.58	32.43	32.58	32.75	**33.28**
	$\times 4$	28.42	30.28	30.14	30.31	30.48	**31.08**
Set14	$\times 2$	30.24	32.28	32.26	32.22	32.42	**32.60**
	$\times 3$	27.55	29.13	29.05	29.16	29.28	**29.55**
	$\times 4$	26.00	27.32	27.24	27.40	27.49	**27.78**
B100	$\times 2$	29.56	31.21	31.16	31.18	31.36	**31.59**
	$\times 3$	27.21	28.29	28.22	28.29	28.41	**28.67**
	$\times 4$	25.96	26.82	26.75	26.84	26.90	**27.15**
Urban100	$\times 2$	26.88	29.20	29.11	29.54	29.50	**30.36**
	$\times 3$	24.46	26.03	25.86	26.44	26.24	**26.88**
	$\times 4$	23.14	24.32	24.19	24.79	24.52	**24.97**

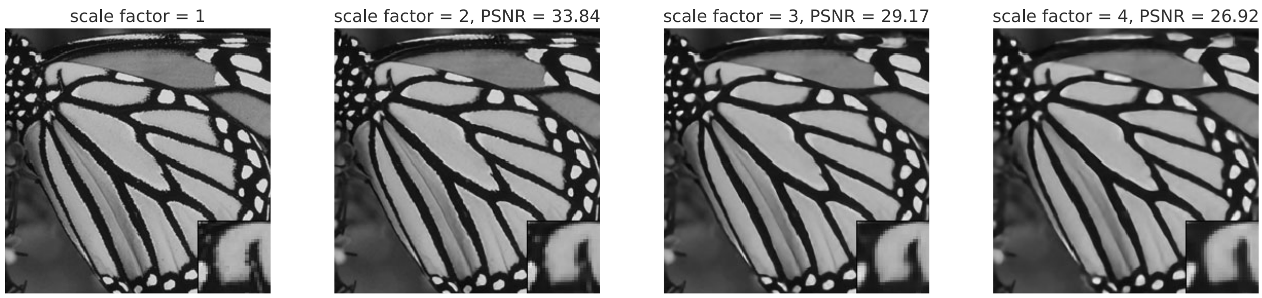

Fig. 3. Example of a low resolution image after applying super-resolution (VDSR method) (original image on the left).

on the original, undistorted images were also more resilient to low image resolution. This was especially the case for top-5 classification accuracy, for which the differences were more significant than in the case of top-1 accuracy.

- For relatively low levels of low resolution, applying super-resolution as a form of image preprocessing allowed us to achieve classification accuracy close to that observed on the original, undistorted images. We therefore conclude that the current state of super-resolution research is sufficient to mitigate the decrease in classification accuracy caused by low levels of low resolution.

- For significantly lowered image resolution, the state-of-the-art super-resolution method considered was still capable of substantially improving classification accuracy. However, the achieved results were far from those observed on undistorted data. Based on the results of our experimental analysis, the existing super-resolution methods are far from being able to completely mitigate the negative impact of very low resolution on classification accuracy.

Several directions for further research can be distinguished. First of all, super-resolution remains an active area of study. It is likely that with future improvements in the quality of super-resolution methods it will be possible to further reduce the negative effect of low resolution on classification accuracy. Especially methods based on generative adversarial networks (Goodfellow *et al*., 2014) seem promising with regard to improving classifiers performance, since they are better suited to producing highly detailed images. It is also possible that achieving the highest possible performance in the restoration task as a super-resolution method does not translate to being the best preprocessing tool in the image recognition task, and designing methods capable of boosting image quality with the goal of improving classification accuracy is necessary.

Furthermore, similar advances are likely to be made in neural architectures used for image recognition. Especially the design of neural networks specifically for dealing with low resolution images could be beneficial. Finally, due to computational constraints, in this paper we did not evaluate the possibility of training the classification network on distorted data. However, based on our previous research on noisy data (Koziarski and Cyganek, 2017), we speculate that this approach could lead to improved classification accuracy at the cost of

Table 2. Classification accuracy depending on the level of resolution degradation and the type of preprocessing used, specifically: baseline accuracy (BLA), classification accuracy on low resolution images (LRA), classification accuracy on low resolution images with the super-resolution applied (SRA), as well as the mitigated performance drop (MPD), calculated as $\mathrm{MPD} = (1 - \frac{\mathrm{BLA-SRA}}{\mathrm{BLA-LRA}}) \times 100\%$. $\mathrm{MPD} = 100\%$ indicates that applying super-resolution allowed complete restoring of baseline performance, and $\mathrm{MPD} = 0\%$ indicates that no improvement was observed due to applying super-resolution.

Network		BLA	Scale	LRA	SRA	BLA - LRA	BLA - SRA	MPD
AlexNet	top-1	54.57	×2	50.69	54.10	3.88	0.47	87.92
			×3	46.09	49.83	8.48	4.73	44.17
			×4	41.41	45.25	13.16	9.32	29.17
	top-5	77.98	×2	74.83	77.62	3.15	0.35	88.76
			×3	70.68	73.83	7.30	4.15	43.17
			×4	66.06	69.69	11.92	8.28	30.52
VGGNet	top-1	69.61	×2	67.29	69.16	2.31	0.45	80.59
			×3	63.92	67.06	5.68	2.55	55.22
			×4	59.29	62.52	10.32	7.08	31.36
	top-5	89.11	×2	87.58	88.86	1.53	0.25	83.46
			×3	85.12	87.36	3.99	1.75	56.04
			×4	81.56	84.37	7.56	4.74	37.30
ResNet	top-1	73.93	×2	71.65	73.50	2.28	0.43	81.15
			×3	67.17	70.65	6.77	3.28	51.51
			×4	61.74	66.11	12.19	7.82	35.84
	top-5	91.47	×2	89.78	91.12	1.69	0.35	79.14
			×3	87.13	89.21	4.34	2.26	47.92
			×4	83.91	86.38	7.56	5.09	32.59

significantly longer training. Evaluating this possibility remains open for further research.

Acknowledgment

This work was supported by the Polish National Science Center under the grant no. 2014/15/B/ST6/00609. The support of the PLGrid infrastructure is also greatly appreciated.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. and Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems, *arXiv:* 1603.04467.

Bevilacqua, M., Roumy, A., Guillemot, C. and Alberi-Morel, M.L. (2012). Low-complexity single-image super-resolution based on nonnegative neighbor embedding, *British Machine Vision Conference (BMVC), Guildford, UK.*

Chang, H., Yeung, D.-Y. and Xiong, Y. (2004). Super-resolution through neighbor embedding, *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, Washington, DC, USA,* Vol. 1, pp. I–I.

da Costa, G.B.P., Contato, W.A., Nazare, T.S., Neto, J.E. and Ponti, M. (2016). An empirical study on the effects of different types of noise in image classification tasks, *arXiv:* 1609.02781.

Dodge, S. and Karam, L. (2016). Understanding how image quality affects deep neural networks, *8th International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal,* pp. 1–6.

Dong, C., Loy, C.C., He, K. and Tang, X. (2014). Learning a deep convolutional network for image super-resolution, *European Conference on Computer Vision, Zurich, Switzerland*, pp. 184–199.

Dong, C., Loy, C.C. and Tang, X. (2016). Accelerating the super-resolution convolutional neural network, *European Conference on Computer Vision, Amsterdam, The Netherlands*, pp. 391–407.

Dutta, A., Veldhuis, R.N. and Spreeuwers, L.J. (2012). The impact of image quality on the performance of face recognition, *33rd WIC Symposium on Information Theory in the Benelux, Boekelo, The Netherlands*, pp. 141–148.

Freeman, W.T., Jones, T.R. and Pasztor, E.C. (2002). Example-based super-resolution, *IEEE Computer Graphics and Applications* **22**(2): 56–65.

Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*, MIT Press, Cambridge, MA.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014). Generative adversarial nets, *in* Z. Ghahramani *et al.* (Eds.), *Advances in Neural Information Processing Systems*, Curran Associates, Inc., Red Hook, NY, pp. 2672–2680.

He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep residual learning for image recognition, *Proceedings of the IEEE*

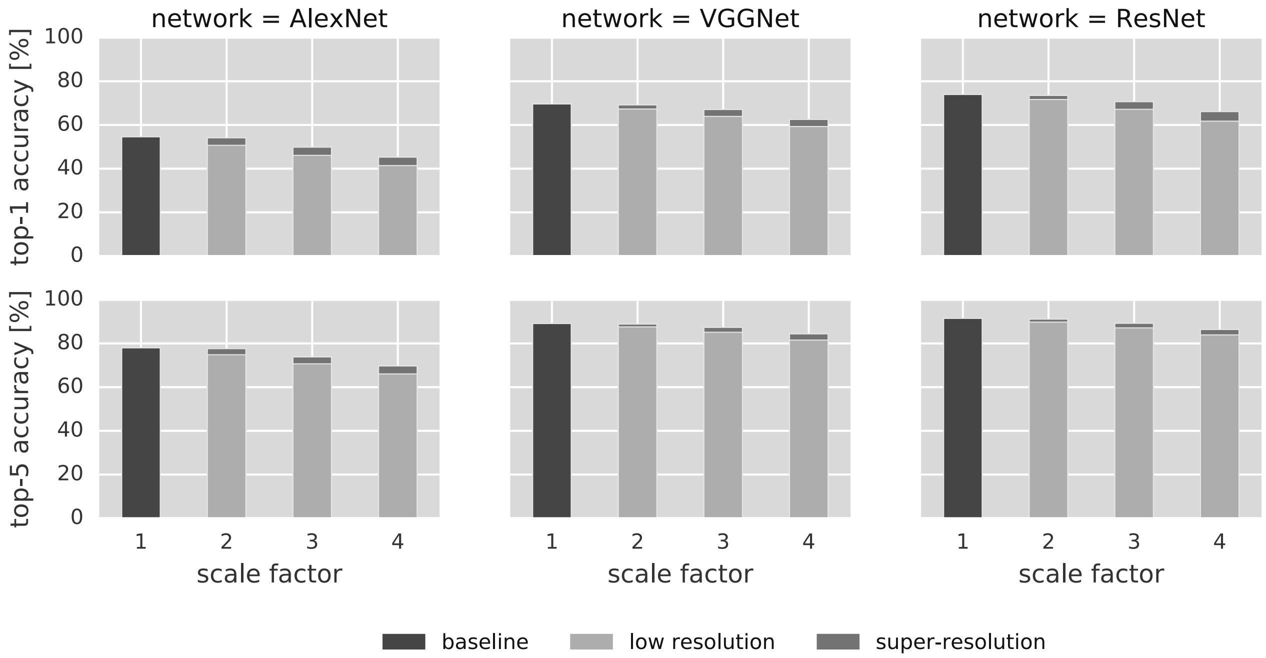

Fig. 4. Impact of low image resolution on classification accuracy, with and without super-resolution applied as a form of preprocessing.

Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 770–778.

Huang, J.-B., Singh, A. and Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA*, pp. 5197–5206.

Johnson, J., Karpathy, A. and Fei-Fei, L. (2016). DenseCap: Fully convolutional localization networks for dense captioning, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, VN, USA*, pp. 4565–4574.

Karahan, S., Yildirum, M.K., Kirtac, K., Rende, F.S., Butun, G. and Ekenel, H.K. (2016). How image degradations affect deep CNN-based face recognition, *International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany*, pp. 1–5.

Kim, J., Kwon Lee, J. and Mu Lee, K. (2016). Accurate image super-resolution using very deep convolutional networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA*, pp. 1646–1654.

Kingma, D. and Ba, J. (2014). Adam: A method for stochastic optimization, *arXiv:* 1412.6980.

Koziarski, M. and Cyganek, B. (2017). Image recognition with deep neural networks in presence of noise—dealing with and taking advantage of distortions, *Integrated Computer-Aided Engineering* **24**(4): 337–349.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks, *Neural Information Processing Systems, Lake Tahoe, CA, USA*, pp. 1097–1105.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W. (2016). Photo-realistic single image super-resolution using a generative adversarial network, *arXiv:* 1609.04802.

Lim, B., Son, S., Kim, H., Nah, S. and Lee, K.M. (2017). Enhanced deep residual networks for single image super-resolution, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshop, Honolulu, HI, USA*, Vol. 1, p. 3.

Long, J., Shelhamer, E. and Darrell, T. (2015). Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA*, pp. 3431–3440.

Mao, X., Shen, C. and Yang, Y.-B. (2016). Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections, *Neural Information Processing Systems, Barcelona, Spain*, pp. 2802–2810.

Martin, D., Fowlkes, C., Tal, D. and Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, *8th IEEE International Conference on Computer Vision, ICCV 2001, Vancouver, Canada*, Vol. 2, pp. 416–423.

Peng, X., Hoffman, J., Stella, X.Y. and Saenko, K. (2016). Fine-to-coarse knowledge transfer for low-res image classification, *IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA*, pp. 3683–3687.

Ronneberger, O., Fischer, P. and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany*, pp. 234–241.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C. and Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge, *International Journal of Computer Vision* **115**(3): 211–252.

Sanchez, A., Moreno, A.B., Velez, D. and Vélez, J.F. (2016). Analyzing the influence of contrast in large-scale recognition of natural images, *Integrated Computer-Aided Engineering* **23**(3): 221–235.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview, *Neural Networks* **61**(1): 85–117.

Schulter, S., Leistner, C. and Bischof, H. (2015). Fast and accurate image upscaling with super-resolution forests, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA*, pp. 3791–3799.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D. and Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA*, pp. 1874–1883.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition, *arXiv:* 1409.1556.

Sun, Y., Wang, X. and Tang, X. (2013). Deep convolutional network cascade for facial point detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA*, pp. 3476–3483.

Tai, Y., Yang, J. and Liu, X. (2017). Image super-resolution via deep recursive residual network, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA*, pp. 2790–2798.

Timofte, R., De Smet, V. and Van Gool, L. (2014). A+: Adjusted anchored neighborhood regression for fast super-resolution, *Asian Conference on Computer Vision, Singapore, Singapore*, pp. 111–126.

Vasiljevic, I., Chakrabarti, A. and Shakhnarovich, G. (2016). Examining the impact of blur on recognition by convolutional networks, *arXiv:* 1611.05760.

Yang, J., Wright, J., Huang, T.S. and Ma, Y. (2010). Image super-resolution via sparse representation, *IEEE Transactions on Image Processing* **19**(11): 2861–2873.

Zeyde, R., Elad, M. and Protter, M. (2010). On single image scale-up using sparse-representations, *International Conference on Curves and Surfaces, Paris, France*, pp. 711–730.

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

Michał Koziarski received his MSc degree in computer science from the Wrocław University of Science and Technology, Poland, in 2016. Currently, he is a PhD student at the Department of Electronics of the AGH University of Science and Technology, Poland. His research interests include computer vision, neural networks and imbalanced data classification.

Bogusław Cyganek received his MSc degree in electronics in 1993, and then his MSc in computer science in 1996, both from the AGH University of Science and Technology, Kraków, Poland. He obtained his PhD degree *cum laude* in 2001 with a thesis on correlation of stereo images, his DSc degree in 2011 with a thesis on methods and algorithms of object recognition in digital images, and a professorship in 2017. In recent years Prof. Bogusław Cyganek has cooperated with many scientific and industrial partners such as Glasgow University, DLR Germany, Surrey University, as well as the Wrocław University of Science and Technology. Currently he is a researcher and a lecturer at the Department of Electronics, AGH University of Science and Technology. His research interests include computer vision, pattern recognition, as well as development of programmable devices and embedded systems. He is an author or a co-author of more than 100 journal and conference papers, as well as several books, with the latest, *Object Detection and Recognition in Digital Images: Theory and Practice*, published by Wiley in 2013. Prof. Cyganek is a member of the IEEE, IAPR and SPIE.

Received: 5 December 2017
Revised: 19 March 2018
Accepted: 12 April 2018