No Need for Glasses or How Blurred Context Can be Helpful for Object Recognition

Anonymous Author(s)
Affiliation
Address
email

Abstract

In this project, we aim to study how blurring the contextual information in images can aid the performance of machine learning models on the task of object recognition. To do that, we trained Faster R-CNN on a set of images with different levels of contextual blurriness - from clear context to highly blurred and completely uninformative context. Then, we tested the models on out-of-context images to assess how they perform in different contextual conditions. Our results indicate that blurring the contextual information in images might be a very simple, but highly effective data preprocessing strategy for achieving higher performance and better robustness.

1 Introduction

We rarely have to recognize a single object on a neutral uninformative background. Usually, objects appear in context – they are often placed in a highly stereotypical backgrounds and they often cooccur with other objects. Previous research has shown that such object cooccurrences and associations between context and objects affect object recognition in both humans and computational models [1, 2, 3, 4].

More specifically, previous studies have shown that context can both help and hurt recognition [3, 4]. Models that make use of context, often perform better when objects appear in typical locations [4]. However, these same models can also perform much worse than usual when presented with out-of-context images [2].

Various solutions have been proposed in the past, to rectify the discrepancy between in-context and out-of-context recognition. Some models have been designed to use only a cropped image of the target object and thus they do not use any contextual information [5, 6]. Some systems have also been designed to use statistical tools to try to decorrelate the object from its background [7]. Other models use attention mechanisms which allow them to process the context separately from the target object [3, 4].

While all of these methods have worked to varying degrees, in this paper, we propose a new simple way of achieving better robustness – by blurring the context information in training data. The advantages of this method are that it is simple and can be applied to any model that uses contextual information.

2 Related Work and Overview

Some effects of blurred context have been studied before by Zhang et al. [3] who trained a computational model on the MSCOCO dataset [8] and observed its performance on a test data with blurred context. Zhang et al.'s original goal was to explore how much blurring is necessary in order to disrupt contextual facilitation in both humans and computational models. However, they found that moderate blurring not only did not reduce the model's accuracy, but also increased it. While they did not discuss the increase of accuracy in their paper, we were interested in studying why that occurred and how we can use it as a tool in object recognition systems.

Therefore, in this project, we aim to study whether object recognition models can improve their performance by preprocessing their training data and blurring the context around the target object (see Figs 1-3). Previous research [3] indicates that removing some contextual information in an image, can be helpful for object recognition in both humans and computational models. At the same time, it has been shown multiple times that context can also be very helpful for object recognition [1, 3, 4]. Inspired by these findings and by the biology of the human eye which exhibits blurred context, we wanted to explore whether blurring can be used as a tool that would allow computational models to concentrate mainly on the target object while still extracting useful information about its surroundings.

We make the following hypotheses:

1. Object recognition models trained on moderately blurred context (Fig. 2) will perform better than object recognition models trained on clear images (Fig. 1).
2. Models trained on moderately blurred context will also exhibit improvement in performance when objects are placed in atypical context.
3. Models trained on images with extremely blurred context (Fig. 3) will perform worse than other models.

To test these hypotheses, we did the following:

1. We created an image dataset where the context in each image is blurred by some factor (Figs 1-3).
2. We implemented and trained Faster R-CNN on every condition in our dataset (for a total of 8 trained models).
3. We tested all of the Faster R-CNN networks on out-of-context images taken from [3] in two separate experiments - one conducted on test images with clear context and another one of test images with blurred context.

All of these experiments helped us quantify the effects of blurred context on the object recognition models and allowed us to study blurring as a potential data preprocessing tool.

3 Data

3.1 Training Data

In order to train and test the performance of various machine learning models, we decided to use the MSCOCO dataset [8]. Consistent with [3], we used 55 categories from MSCOCO and we created 8 conditions based on the blurriness level of the context. To blur the images, we used Gaussian blur with radius $r \in \{0, 1, 2, 4, 8, 16, 32, 64\}$ (Figs 1-3).

The final data consisted of 97,069 training images in each condition, for a total of 776,552 images across all conditions.

3.2 Testing Data

To test the performance of the model, we used the data from [3]. Their cut-and-paste dataset consists of out-of-context images grouped in 4 contextual conditions and 4 bins based on the size of the target object for a total of 16 groups.

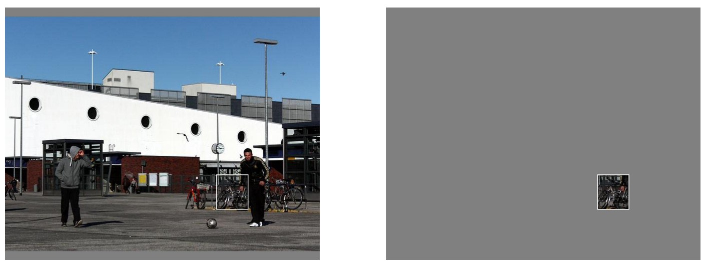
The data contains the following contextual conditions - full context, minimal context, congruent context, and incongruent context. Full context represents the object in a natural setting. Minimal context is just the target object cropped around its bounding box, so that there is very little contextual information. Congruent context images consist of a cut-and-pasted target object in a different image that contains an object with the same label. Analogously, incongruent context images consist of a

Figure 1: The original image that contains no blurring

Figure 2: Gaussian blurring with radius 8 is applied around the target object.

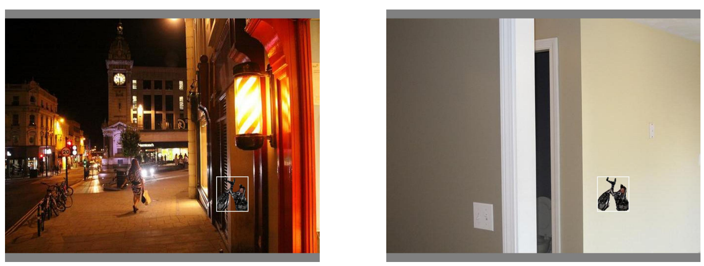
Figure 3: Gaussian blurring with radius 32 is applied around the target object.

cut-and-pasted target object in a different image that does not contains an object with the same label so that the context is more unnatural. For examples of all of these conditions, see Figure 4.



a) full context

b) minimal context



c) congruent context

d) incongruent context

Figure 4: **Examples of the Cut-and-paste Dataset**. The same target object - bike, placed in the four contextual conditions from the cut-and-paste dataset. Image generated by me but originally used in a different paper.

In this project, we used 7,772 cut-and-paste images across all 16 conditions. We conducted two experiments - in one of them, we tested the computational model on the original images with no blurred context, and in the other experiment, we blurred the context before testing (using the same blur radius as the one used when training the model). This data allowed us to test how the models generalizes to different contextual conditions as a function of the blur radius.

4 Computational Model, Code, and Supplements

To test how blurring affects recognition, we used an implementation of Faster R-CNN originally used in [4]. The training and testing images were taken from [3] but were modified for the purposes of this project by us. All of the code that we have used and written can be found at `https://github.com/DKarev/blurred-context`. For detailed results from testing and training, also refer to the github repo.

5 Results

In this section, we will present the most important results from our experiments. For all results, please refer to `https://github.com/DKarev/blurred-context`.

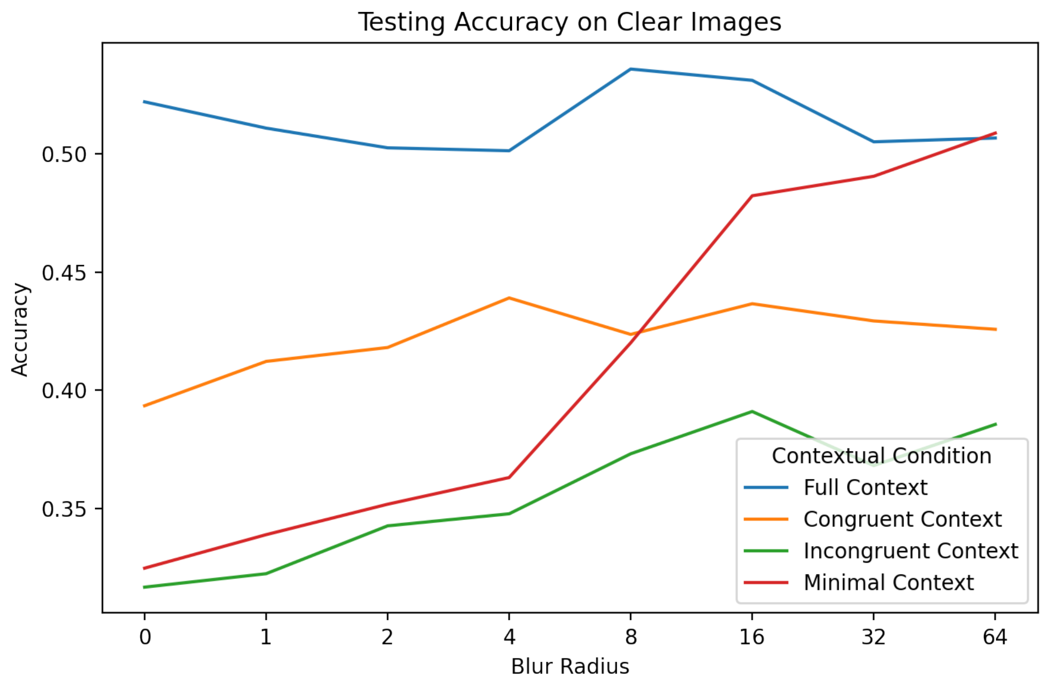


Figure 5: **Accuracy of the models trained on images with blurred context and tested on images with clear context**. Results averaged over all object size conditions. The blur radius of the context used for training (but not testing) is on the x-axis, whereas the accuracy for each network in every condition is represented on the y-axis.

The first experiment that we conducted was testing each of the trained neural networks on images with clear context (blur radius equal to 0). The results of that experiment are summarized in Figure 5. The second experiment that we conducted was testing the same neural networks on images with blurred context (both training and testing data generated with the same blur radius). The results of that experiment are summarized in Figure 6.

We can notice the following facts from the results:

1. **Training on images with blurred context can improve the performance of the model in all contextual conditions.** We can see that the performance of Faster R-CNN increases as we blur the context in the training data. This is true for all contextual conditions to certain extend. We hypothesize that that is the case due to the model relying less on contextual information and more on information about the object.
2. **Training on blurred context produces best results when the context in the images is uninformative or confusing.** Even though the performance of the model on both the full context condition and the congruent context condition has improved to some degree, the highest improvement is observed in the minimal and incongruent conditions. This is expected, since relying on context in the full context condition and in the congruent context condition is not necessarily a bad strategy.
3. **For most conditions, the accuracy of the model peaks for intermediate values of the blur radius.** The minimal context condition is the only condition for which increasing the blur radius always improves the performance of the model and this is expected since it is also the only condition that does not use any contextual information. The performance for the other contextual conditions is reduced for higher values of the blur radius, since the

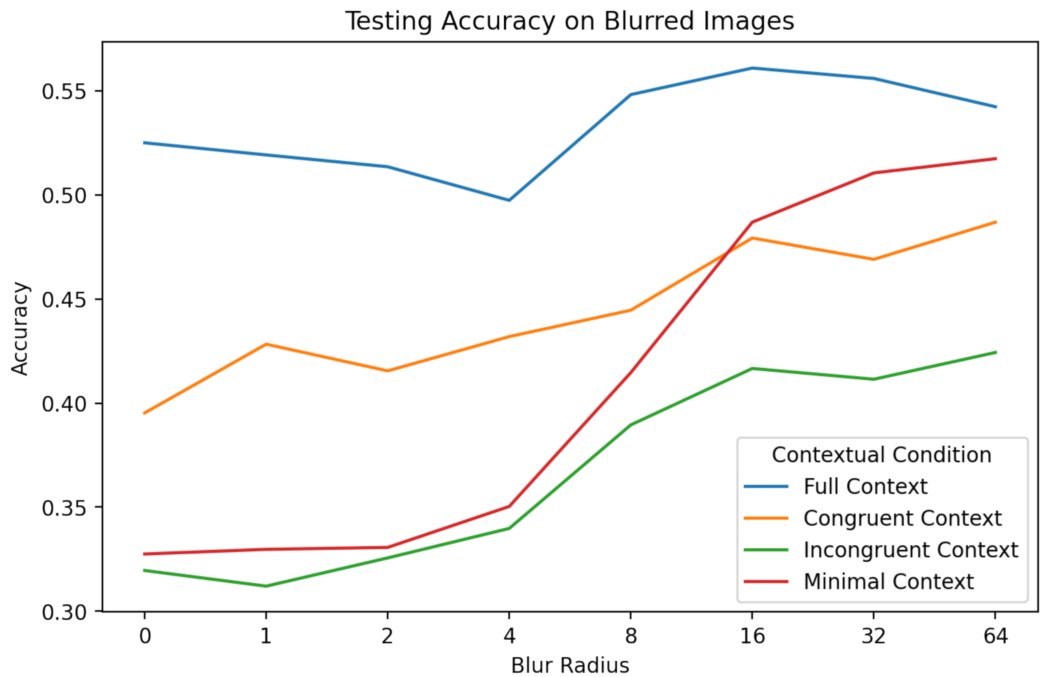


Figure 6: **Accuracy of the models trained on images with blurred context and also tested on images with the same level of blurred context**. Results averaged over all object size conditions. The blur radius of the context used for training (and testing) is on the x-axis, whereas the accuracy for each network in every condition is represented on the y-axis.

model is forced to ignore all contextual information which can be detrimental. It seems like the optimal level of blurriness is when the blur radius is around 16.

4. **Accuracy increases when the testing data also has blurred context.** When we compare Figure 5 and Figure 6, we can see that when also tested on images with blurred context, the accuracy of the models increases.

6 Discussion and Future Work

We created 8 different versions of the MSCOCO dataset [8] based on the level of blurriness of the context varied from 0 (i.e. clear context) to 64 (i.e. very blurred context). We trained Faster R-CNN networks on each version of the data and then we tested each network on the cut-and-paste data provided by [3]. For each network, we conducted two experiments - (1) testing it on images with clear context and (2) testing it on images with context that is as blurred as its training set.

We showed that an incredibly simple preprocessing strategy such as blurring can improve model's generalizability and performance on a variety of contextual conditions, including images containing beneficial contextual information (i.e. full context and congruent context). While the observed effects of blurring can be relatively weak for some conditions, the advantage of using this technique are in its simplicity (i.e. blurring images takes up very little processing time and can be done very easily) and its wide applicability (i.e. blurring can be used on any image and in any model that processes some contextual information).

Here, we would like to emphasize the following big picture conclusions that our results support:

1. **Neural networks can benefit from reduced information in their training set.** While recently, there has been a lot of efforts in the field of object recognition to obtain more data [9], our results indicate that more data is not necessarily needed for achieving higher performance. We showed that performance can also be increased by using carefully preprocessed data that guides the neural network towards higher generalizability.

2. **Context can both help and hurt recognition.** Similarly to results from previous studies [3] [4], we find that contextual information in images can be both helpful and hurtful. Neural networks that use context such as Faster R-CNN will rely to certain extend on contextual information. If the images that they are recognizing have helpful context, then that dependence on context will be useful. However, if the images are out-of-context, then that dependence might hurt the networks' performance.
3. **Blurring can play an important role balancing networks' dependence on context.** Our results indicate that blurring can help models generalize better by still using context but relying less on it.

While, in general our results indicate that blurring can be a beneficial preprocessing step for object recognition, it is also worth noting some of the limitations of the current study and possible directions for future work.

1. **Training and testing other computational models.** Due to limited time and computational resources, we were able to only train and test Faster R-CNN. To confirm the results outlined in this project, future research might need to repeat our methodology on other computational models that use contextual information. One possible candidate for this would be VGG-16 [10]. Testing on other computational models, would ensure that our results are not specific to Faster R-CNN.
2. **Comparing performance to models trained on images containing no contextual information.** While we believe that it is unlikely that the results we obtained can be improved by completely ignoring contextual information, it would be an interesting point of comparison. Blurring seems to play a balancing role in how much context a model would use, but we still do not know the exact consequences on performance of completely removing contextual information.
3. **Design better blurring algorithms.** One interesting direction for future research might be designing a better way of blurring images. We know that some parts of the context are more important than others (e.g. the contextual information right around the object versus the contextual information in the corners of the image). However, the blurring algorithm that we used treats all contextual information equally and thus, a different algorithm that acts preferentially might achieve better results.

References

[1] A. Oliva and A. Torralba, "The role of context in object recognition," *Trends in cognitive sciences*, vol. 11, no. 12, pp. 520–527, 2007.

[2] A. Rosenfeld, R. Zemel, and J. K. Tsotsos, "The elephant in the room," *arXiv preprint arXiv:1808.03305*, 2018.

[3] M. Zhang, C. Tseng, and G. Kreiman, "Putting visual object recognition in context," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 985–12 994.

[4] P. Bomatter, M. Zhang, D. Karev, S. Madan, C. Tseng, and G. Kreiman, "When pigs fly: Contextual reasoning in synthetic and natural scenes," *arXiv preprint arXiv:2104.02215*, 2021.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[6] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[7] K. K. Singh, D. Mahajan, K. Grauman, Y. J. Lee, M. Feiszli, and D. Ghadiyaram, "Don't judge an object by its context: Learning to overcome contextual bias," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 070–11 078.

[8] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[9] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.