MODELING HUMAN DEVELOPMENT: EFFECTS OF BLURRED VISION ON CATEGORY LEARNING IN CNNS

Anonymous authors
Paper under double-blind review

ABSTRACT

Recently, training convolutional neural networks (CNNs) using blurry images has been identified as a potential means to produce more robust models for facial recognition (Vogelsang et al., 2018). This method of training is intended to mimic biological visual development, as human visual acuity develops from near-blindness to almost normal acuity in the first three to four months of life (Kugelberg, 1992). Object recognition develops in tandem during this time, and this developmental period has been shown to be critical for many visual tasks in later childhood and adulthood. We explore the effects of training CNNs on images with different levels of applied blur, including training regimens with progressively less blurry training sets. Using subsets of ImageNet (Russakovsky et al., 2015), CNN performance is evaluated for both broad object recognition and fine-grained classification tasks. Results for AlexNet (Krizhevsky et al., 2012) and the more compact SqueezeNet (Iandola et al., 2016) are compared. Using blurry images for training on their own or as part of a training sequence increases classification accuracy across collections of images with different resolutions. At the same time, blurry training data causes little change to training convergence time and false positive classification certainty. Our findings support the utility of learning from sequences of blurry images for more robust image recognition.

1 INTRODUCTION AND RELATED WORK

In the first three to four months of life, human visual acuity develops quickly from about 20/800 to 20/40 (Kugelberg, 1992). This developmental period is important to perceptive and visual tasks later in life such as peripheral vision, global motion, and facial recognition (Lewis & Maurer, 2005; De Heering et al., 2012). In fact this period of blurry vision seems to be particularly critical to the development of facial recognition in childhood and later in life (Röder et al., 2013).

Convolutional neural networks (CNNs) are the current state of the art in many computer vision applications, and have come to be the dominant model for object recognition. Currently, the most accurate CNNs exceed human performance for object recognition under certain specific testing conditions (Geirhos et al., 2018). However they are not nearly as robust as the human visual system, and tend to suffer disproportionately when trying to classify degraded or blurry images (Karahan et al., 2016). Blurry images, degraded images, and images with object occlusion are common in many computer vision applications, and being able to deal with these images is important to having a robust system.

Vogelsang et al. have investigated new possible methods of training CNNs specifically for facial recognition by using training sets consisting of deliberately blurred images (Vogelsang et al., 2018). They report an increase in facial recognition accuracy when training the CNN AlexNet on blurry images and then clear ones, as opposed to training on clear images only. They hypothesize that the poor initial visual acuity experienced by newborns helps them develop larger receptive fields, which are instrumental in spatial vision, and that this benefit can be achieved in CNNs as well. However, Katzhendler and Weinshall dispute the conclusions drawn by Vogelsang et al., saying that the original methodology was flawed and in fact their own simulations with artificial networks do not support the hypothesis that high initial visual acuity is detrimental to object recognition (Katzhendler & Weinshall, 2019). Katzhendler and Weinshall claim that the only reason Vogelsang et al. found a benefit to training with blurred images is because they also included blurred images in their test

set, and they go on to demonstrate that including blurred images in training is actually detrimental to identifying only clear images. Their point is that training with blur will not provide a benefit in general because the network will already have captured important low-frequency features when training on clear images.

In our investigation we broaden the study of the benefits and drawbacks of learning with image blur to investigate general varied object recognition beyond face specialization, and compare these results to fine-grained object class identification. We further investigate the effect of blurred training images on more compact network architectures, overall training time required for convergence, and the uncertainty associated with viewing objects from no known class. We use two image sets each with five different levels of blur, and study two popular convolutional networks, AlexNet and SqueezeNet. We also test sequences of blurry images of differing lengths. We claim that the results of Vogelsang et al. and Katzhendler and Weinshall are not contradictory, and indeed are totally compatible. We show that while it is true that using blurred images in training causes a drop in accuracy when identifying higher resolution images, the reciprocal increase in accuracy on the set of images of all blur levels as a whole is almost always disproportionately larger. Additionally, there is no significant difference in required training time when using blurry images and the certainty of misclassified images is not significantly different. Therefore if the goal is robust real-world object identification, using blurred training images is ultimately beneficial.

2 METHODS

2.1 DATASETS

We focus on two datasets, Imagenette and Imagewoof (Howard, 2020). These image sets are two subsets of the well-known dataset ImageNet (Russakovsky et al., 2015) intended for fast training and testing. Imagewoof is a set of ten classes, all of which are breeds of dogs. Of those ten, we selected seven classes to use for training in order to speed up the training process. Imagenette is a set of ten visually distinct object classes, meant for quick training to distinguish objects which share very few similarities. Similarly, we selected seven of those classes to use for training.

These two datasets were selected because Imagewoof contains object classes that tend to be visually similar to one another as they are all dog breeds, while Imagenette contains object classes that are very visually distinct. This is so we can compare the effect of having blurry training images when discrimination is required between classes that share many visual features with discrimination between classes that do not. For both datasets, there are 1300 images belonging to each class. We train the CNNs AlexNet and SqueezeNet to classify these images.

2.2 NETWORK ARCHITECTURES

We compare the results of using blurry training data on the two CNNs AlexNet and SqueezeNet. AlexNet is a popular reference network in computational neuroscience due to its accuracy and the fact it is well studied. SqueezeNet was chosen as a comparison because it is much smaller then AlexNet in terms of parameters but achieves a similar accuracy under circumstances of normal training (Iandola et al., 2016) . AlexNet is a standard convolutional network with about 61 millions parameters (Krizhevsky et al., 2012). SqueezeNet is a deeper network with makes use of Fire modules which squeeze and expand the input, ultimately requiring many fewer parameters than AlexNet, only about 1.2 million (Iandola et al., 2016). We implement the models in PyTorch.

2.3 IMAGE BLURRING

Gaussian blurring consists of convolving an image with a Gaussian kernel. The convolving kernel contains values distributed according to the two-dimensional Gaussian function. We compute the standard deviation to use in the Gaussian function from the size of the kernel. An image blurred using a Gaussian function with a kernel size of one pixel is approximately equivalent to a clear image.

To a rough approximation, the objects in the images we intend to classify would subtend a visual angle of 5 - 10 degrees for an average observer, so the entire image may subtend a 5 - 20 degree

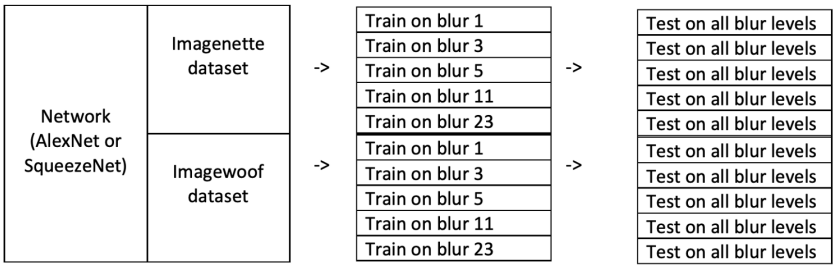

Figure 1: Diagram of training and testing flow for single blur training

visual angle. AlexNet and SqueezeNet accept input images of size 227 x 227 x 3 pixels (Krizhevsky et al., 2012). At the age of one month, infant visual acuity is about 30 arc minutes or 1/2 of a degree. Given the size of the input images, 1/2 of a degree of visual angle would correspond to about 23 pixels (high estimate). Therefore for the highest blur we used a square Gaussian kernel with a side length of 23 pixels. A similar analysis was repeated for the average childhood visual acuities at increasing ages, up to eight months old. For Gaussian blurring each kernel is required to have a side length that is a positive odd integer. Given this constraint the ultimate set of blur levels we use consists of square Gaussian kernels of side length one pixel, three pixels, five pixels, eleven pixels, and twenty-three pixels (going from least blurry to most blurry).

2.4 Network Training

We tested two training methods, single blur and sequential blur training. During single blur training, each network is trained using images that are all equally blurry and then tested on images of all blur levels. During training, 5-fold cross validation is used. We also record the network weights at select points for additional analysis. All single blur training was done for 100 epochs only, in accordance with the time and computational constraints of our lab. An outline of this training regime is visualized in Fig 1.

During sequential blur training we train each network on varying sequences of images with different levels of blur. For example, we first train the network on images with a blur window of one pixel only. Then we train on blur of three pixels followed by blur of one pixel and test again. Then we train on five pixel blur, followed by three pixel blur, followed by one pixel blur, and test again. We repeat this process until we include all blur levels in the sequential training. At test time we test on images of all blur levels. We completed two types of sequential blur training. The first we called long training, where for a given sequence of blur levels we trained the network on each blur level for 100 epochs. The second, short training, consisted of training the network for a total of 200 epochs, split evenly between however many blur levels were in a given sequence.

For both AlexNet and SqueezeNet we use a batch size of 128. For training we use a stochastic gradient descent optimizer with a learning rate of 0.001, momentum of 0.9, and weight decay of 0.0005 as originally recommended by Krogh and Hertz (Krogh & Hertz, 1992). 100 epochs of training takes between six and eight hours each run for single blur training, and sixteen to twenty hours for sequential blur training. In both networks, the dense (linear) layers are initialized using a random uniform distribution with standard deviation equal to the reciprocal of the square root of the number of weights, and the convolutional layers are initialized using Kaiming initialization (He et al., 2015).

3 Results

3.1 Single Blur Training

We find that our results generally support the idea that using blurred images in training can increase accuracy at test time and also makes the network more robust. This holds to some extent for both network architectures and both image sets that we tested. Looking first at the results of training

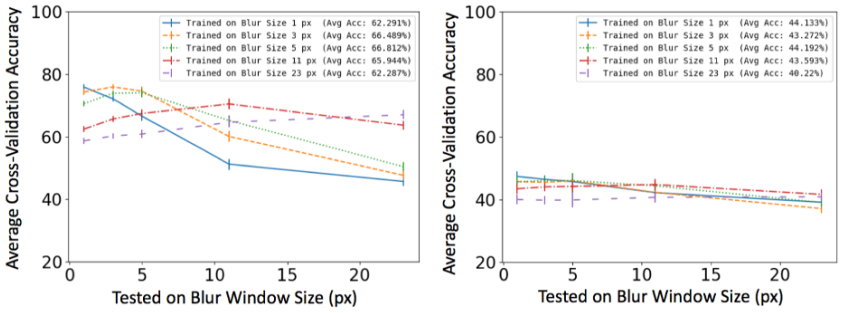

Figure 2: Average cross-validation accuracy of AlexNet trained on each level of blur and tested on all five blur levels using selected images from Imagenette (left) and Imagewoof (right).

AlexNet — Imagenette — Test Blur

Training Blur	1	3	5	11	23	Total Avg
1	75.81	72.08	66.56	51.25	45.76	62.29
3	74.26	75.86	74.53	60.07	47.73	66.49
5	70.53	73.85	74.02	65.19	50.47	66.81
11	62.44	65.68	67.44	70.44	63.72	65.94
23	58.67	60.24	60.86	64.64	67.03	62.29

AlexNet — Imagewoof — Test Blur

Training Blur	1	3	5	11	23	Total Avg
1	47.32	46.39	45.68	42.18	39.10	44.13
3	45.67	45.46	45.93	42.26	37.04	43.27
5	45.56	46.05	45.98	44.35	39.03	44.19
11	43.42	44.04	44.15	44.74	41.63	43.59
23	39.99	39.77	39.82	40.68	40.84	40.22

SqueezeNet — Imagenette — Test Blur

Training Blur	1	3	5	11	23	Total Avg
1	77.02	65.17	57.81	47.03	40.59	57.52
3	59.20	74.40	73.30	54.79	45.08	61.36
5	49.10	68.91	72.83	60.54	47.98	59.87
11	50.12	60.29	65.34	71.21	55.85	60.56
23	52.83	57.60	60.37	66.15	69.47	61.28

SqueezeNet — Imagewoof — Test Blur

Training Blur	1	3	5	11	23	Total Avg
1	54.04	42.34	37.39	31.01	29.81	38.92
3	45.86	49.61	47.89	40.73	32.70	43.36
5	43.47	49.53	49.51	43.34	35.24	44.22
11	34.12	38.43	40.94	46.70	39.61	39.96
23	30.58	33.03	34.51	40.09	45.08	36.66

Figure 3: Single blur training accuracies at different training and testing blur levels, including averages over testing blur levels.

AlexNet using the Imagenette dataset we find that the highest mean cross-validation accuracy always occurs when testing the network on images with the same level of blur as the images the network was trained on, which can be seen in Fig 2. The corresponding numerical values are recorded in Fig 3. This result is as we would expect. The highest accuracy at any one test blur is found when classifying the clearest images after training on high resolution images, about 76%, and in general the accuracy decreases the farther away the training and testing blur levels are from each other. The lowest accuracy at any one test blur is found when classifying the blurriest images after training on higher resolution images, about 46%. This trend is seen in the results for the SqueezeNet architecture as well, in Fig 4.

Compared to AlexNet, the overall performance of SqueezeNet is slightly worse on both image sets. The disparity between training on different levels of blur is clearer though, as can be seen in Fig 4 where the testing curves are significantly sharper than those belonging to AlexNet in Fig

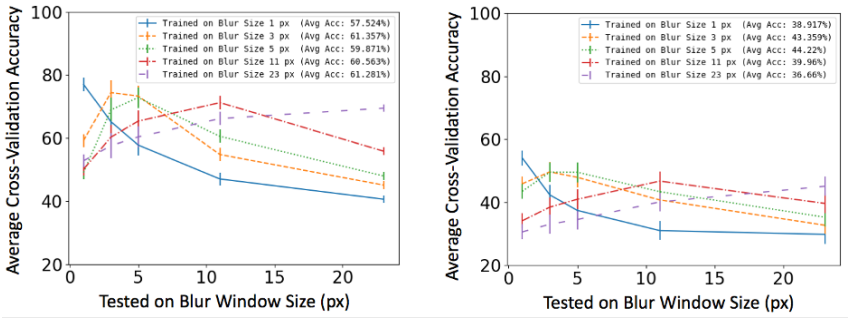

Figure 4: Average cross-validation accuracy of SqueezeNet trained and tested using selected images from Imagenette (left) and Imagewoof (right).

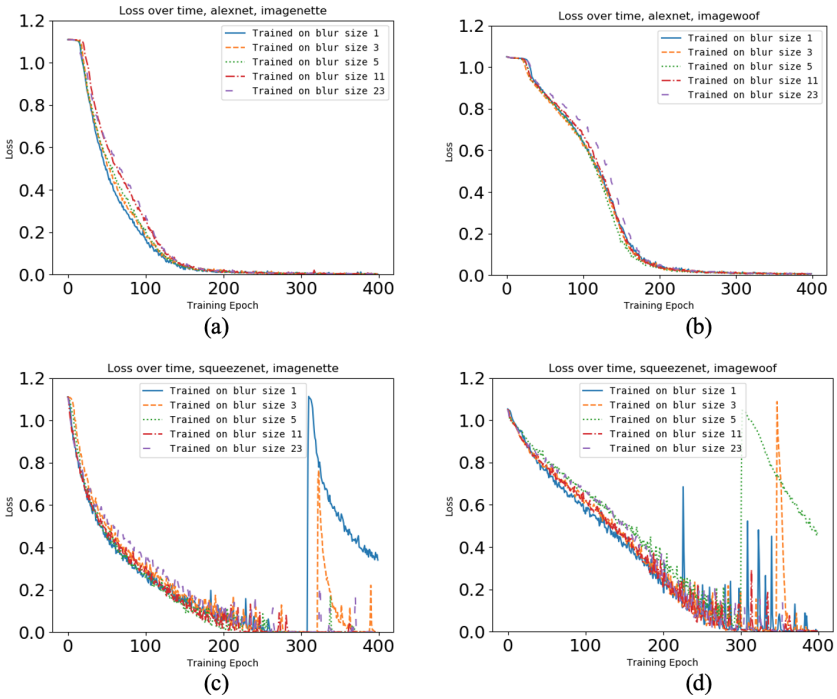

Figure 5: Training loss at each epoch: (a) AlexNet trained on Imagenette, (b) AlexNet trained on Imagewoof, (c) SqueezeNet trained on Imagenette, (d) SqueezeNet trained on Imagewoof

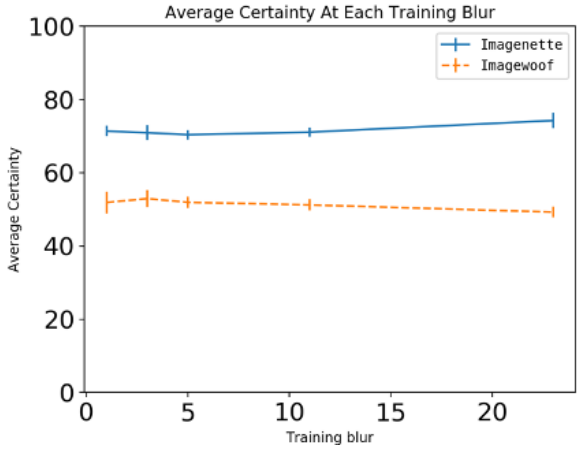

Figure 6: AlexNet?s reported certainty when attempting to classify new images that don?t belong to any known class.

2. SqueezeNet appears to be more sensitive to varying levels of training blur. We hypothesize that this is due to the fact that SqueezeNet is so much smaller than AlexNet and has fewer parameters to work with. Having fewer parameters may make it more difficult for a network to generalize to statistically divergent test data.

We also look at the convergence of the loss over time during training on sets of images at different blur levels, shown in Fig 5. We observe that in AlexNet, training losses on different blur levels diverge in the middle of training, but all end up converging around the same time at the end of training. In SqueezeNet there is more noise so the trend is less clear. There also appear to be some numerical artifacts in SqueezeNet training at later epochs.

Ultimately we find that while training on high-blur images leads to less accurate recognition of low-blur images and vice versa, the gaps in performance notably differ. The important thing we note is that learning from lower-resolution images allows somewhat strong recognition for higher-resolution images, while learning from higher-resolution images does not equivalently benefit lower-resolution recognition. Looking at the average accuracies for classifiers trained on each blur level and each image set, we find that the highest accuracy will be achieved across all image resolutions by using a training set of images with blur windows of either size 3 pixels or 5 pixels, rather than higher resolution images, except in the case of training AlexNet on Imagewoof where the result is unclear within the bounds of error.

Lastly we also investigate what we call false positive classification certainty. This is the certainty that the network reports when it is classifying an image that does not belong to any class it has seen before. Such an image cannot be accurately classified, and so in the ideal case a robust network should report low certainty when identifying what it thinks the class of the image should be. We are interested in whether the network grows more uncertain when classifying unseen images from new classes if the training data is more blurry. Asking AlexNet to classify and image from an object class that it was not trained to recognize, we find that the network?s false certainty doesn?t change significantly according to whether it was trained on blurry images. These results are shown in Fig 6.

3.2 Sequential Blur Training

Next, we look at training on image sets with sequentially higher resolutions, from the lowest resolution images to highest. We test intermediate blur sequences of several different lengths. The results are shown in Fig 7. We find that in all cases, training a network on either all blur levels or some intermediate sequence of blur levels will give the best cumulative performance for all combined testing blur levels. Indeed in every case, a sequence of blur levels of any length in training offers a significant increase compared to training on only high resolution images when it comes to

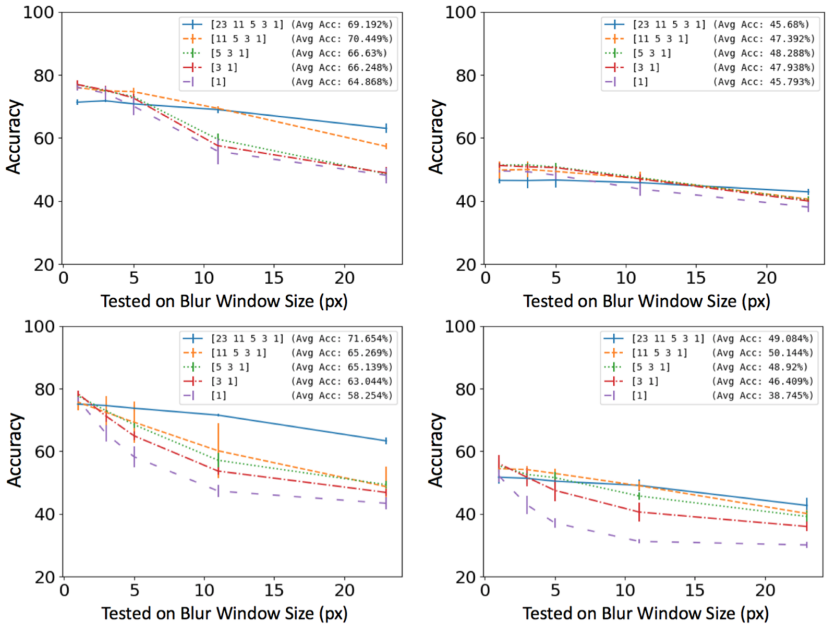

Figure 7: Classification accuracy after sequential blur training: (a) AlexNet trained on Imagenette, (b) AlexNet trained on Imagewoof, (c) SqueezeNet trained on Imagenette, (d) SqueezeNet trained on Imagewoof

test time. Training on a sequence of low to high resolution images allows for significantly more robust image recognition that is impacted less by degradation that comes from image blur. In some cases, this overall gain comes at the cost of losing some accuracy when classifying high resolution images, particularly with AlexNet. For most training regimes, classification accuracy for higher resolution images is close to equivalent for higher blur training, while improvements for higher blur classification are still significant.

We also complete sequential training using what we call short training, as opposed to the long training in Fig 7. During short training we train each network for a total of 200 epochs split evenly between the blur levels that compose the sequence. The results of the uniform-length sequential training are shown in Fig 8. These results confirm what we saw in Fig 7, that overall there is a net benefit of training on a sequence of blur levels to increase robustness in object classification. We note that when testing on images with a blur window of one pixel, we will usually suffer a decrease in accuracy by including any blurry images in the training set. However, testing on images with a blur window of even three or five pixels shows us that having any level of blurred images in the training set is beneficial overall. This provides additional support for the hypothesis that the extended period of visual development in infants can benefit vision tasks for the rest of life, and provides a hint into how we may start to bridge the large gap that still remains between human and computer vision.

4 Conclusion

Our body of experiments and results show that training CNNs with blurred images is sometimes slightly harmful to the identification accuracy of clear images; however when this decrease in accuracy happens, it happens in tandem with a disproportionately larger increase in accuracy at identifying images of different blur levels. In most cases we tested, the average accuracy across all blur levels is greatest when some level of blurry images is included in the training set. Sequential blur training is more effective than single blur, but to ensure robustness to image degradation any blur training was shown to be better than training only on high resolution images. Training on blurry images does not take longer to converge than training on clear images. We also show that false certainty about incorrect classifications does not increase when training is done on sets of blurry images.

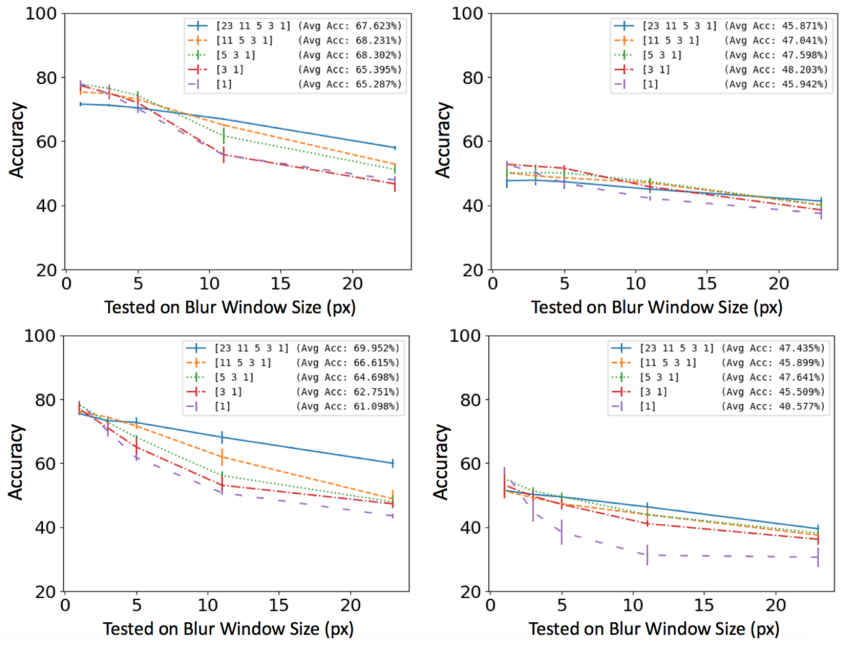

Figure 8: Uniform-length sequential blur training: (a) AlexNet trained on Imagenette, (b) AlexNet trained on Imagewoof, (c) SqueezeNet trained on Imagenette, (d) SqueezeNet trained on Imagewoof

CNNs currently outperform humans at object identification tasks for clear, high quality images, but they cannot compete with human vision when identifying degraded or occluded images. The human visual system is far more robust than CNN vision for reasons that are not yet entirely clear. If our goal is to approach closer to human vision with machine learning, including blurry images in training is one thing that does seem to bring CNNs closer to that goal. This strategy increases robustness without increasing necessary training time or false certainty, while sometimes sacrificing some identification power on high resolution images. In many real-world applications this effect is desirable, such as computational analysis of satellite imagery where images might be degraded by atmospheric conditions . Any ecosystem where blurry images are either equally or more likely to be encountered than clear images will most likely benefit from using blurry images in early training.

It is not the case that training on only high resolution images is the best you can do to generalize a neural network. In principle CNNs should be able to learn low frequency features just as well and at the same time as high frequency features, but there is still a general benefit to learning with blurred training data. Perhaps there is some type of low frequency information that it is beneficial to force your network to learn first before allowing it to train on higher resolution images with more higher frequency features.

The natural next step to follow up this work is to expand this line of inquiry to larger datasets and test on more varied network architectures. We were limited by time and computing power constraints, but this proof of concept shows that it would be worthwhile to devote more resources to investigating this area of research. In the future we also want to fully develop a calculus to decide what specific level of blur to use on training images to best generalize a neural network for some task. As yet it is unclear what the deciding factor is that makes one training blur level outperform another in broad testing. An additional interesting follow up to this work would be to examine other aspects of human vision that develop in the months after birth along with visual acuity (Lewis & Maurer, 2005). These factors include color saturation and color discrimination, and they may play a similar role as visual acuity in training the developing human visual processing system in young infants.

REFERENCES

Adélaïde De Heering, Bruno Rossion, and Daphne Maurer. Developmental changes in face recognition during childhood: Evidence from upright and inverted faces. *Cognitive Development*, 27 (1):17–27, 2012.

Robert Geirhos et al. *Generalisation in humans and deep neural networks*. Advances in neural information processing systems, 2018.

Kaiming He et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, 2015.

Jeremy Howard. Imagenette, 2020. URL `https://github.com/fastai/imagenette`.

Forrest N. Iandola et al. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. preprint, arXiv, 2016.

Samil Karahan et al. How image degradations affect deep cnn-based face recognition? 2016, 2016.

Gal Katzhendler and Daphna Weinshall. Blurred images lead to bad local minima. preprint, arXiv, 2019.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. *Imagenet classification with deep convolutional neural networks*. Advances in neural information processing systems, 2012.

Anders Krogh and John A. Hertz. *A simple weight decay can improve generalization*. Advances in neural information processing systems, 1992.

Ulla Kugelberg. Visual acuity following treatment of bilateral congenital cataracts. *Documenta ophthalmologica*, 82(3):211–215, 1992.

Terri L. Lewis and Daphne Maurer. Multiple sensitive periods in human visual development: evidence from visually deprived children. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 46(3):163–183, 2005.

Brigitte Röder et al. Sensitive periods for the functional specialization of the neural system for human face processing. *Proceedings of the National Academy of Sciences*, 110(42):16760–16765, 2013.

Olga Russakovsky et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

Lukas Vogelsang et al. Potential downside of high initial visual acuity. *Proceedings of the National Academy of Sciences*, 115(44):11333–11338, 2018.