PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Robustness of adversarial camouflage (AC) for naval vessels

Løkken, Kristin Hammarstrøm, Brattli, Alvin, Palm, Hans Christrian, Aurdal, Lars, Klausen, Runhild Aae

Kristin Hammarstrøm Løkken, Alvin Brattli, Hans Christrian Palm, Lars Aurdal, Runhild Aae Klausen, "Robustness of adversarial camouflage (AC) for naval vessels," Proc. SPIE 11394, Automatic Target Recognition XXX, 113940W (24 April 2020); doi: 10.1117/12.2558506

SPIE.

Event: SPIE Defense + Commercial Sensing, 2020, Online Only, California, United States

Robustness of Adversarial Camouflage (AC) for Naval Vessels

Kristin Hammarstrøm Løkken[a], Alvin Brattli[a], Hans Christian Palm[a], Lars Aurdal[b], and Runhild Aae Klausen[a]

[a]Norwegian Defence Research Establishment (FFI), P.O. Box 25, 2027 Kjeller, Norway
[b]Independent researcher

ABSTRACT

Different types of imaging sensors are frequently employed for detection, tracking and classification (DTC) of naval vessels. A number of countermeasure techniques are currently employed against such sensors, and with the advent of ever more sensitive imaging sensors and sophisticated image analysis software, the question becomes what to do in order to render DTC as hard as possible. In recent years, progress in deep learning, has resulted in algorithms for image analysis that often rival human beings in performance. One approach to fool such strategies is the use of adversarial camouflage (AC). Here, the appearance of the vessel we wish to protect is structured in such a way that it confuses the software analyzing the images of the vessel. In our previous work, we added patches of AC to images of frigates. The paches were placed on the hull and/or superstructure of the vessels. The results showed that these patches were highly effective, tricking a previously trained discriminator into classifying the frigates as civilian. In this work we study the robustness and generality of such patches. The patches have been degraded in various ways, and the resulting images fed to the discriminator. As expected, the more the patches are degraded, the harder it becomes to fool the discriminator. Furthermore, we have trained new patch generators, designed to create patches that will withstand such degradations. Our initial results indicate that the robustness of AC patches may be increased by adding degrading filters in the training of the patch generator.

Keywords: naval vessel, camouflage, artificial intelligence, neural network, robustness

1. INTRODUCTION

Different types of imaging systems are frequently employed for detection, tracking or classification of naval vessels. Such systems may include one or more imaging sensors combined with one or more image processing platforms running the required algorithms. A number of countermeasure techniques are currently employed against such imaging systems. Depending on the observation spectrum employed by the sensor system they can be:

- Signature reduction aimed at simply reducing the signal from the vessel.
- Camouflage in the form of shape or colour changing approaches that aim at modifying the appearance of the vessel.
- Flares/decoys that generate smoke screens or generate artificial targets.
- Active countermeasures such as laser illumination aimed at blinding or confusing the sensors.

With the advent of ever more sensitive and sophisticated imaging sensors combined with steadily improving processing platforms running more and more advanced algorithms, it becomes increasingly difficult to avoid detection, tracking and classification. In recent years, progress in neural networks and machine learning, often described as deep learning, has led to a performance leap for image analysis algorithms, empowering image analysis algorithms that outperform previous algorithms to a substantial degree. Knowing that the vessels' visual and infrared signatures cannot be reduced to zero, and assuming that flares and decoys cannot mask

Further author information:
Kristin H. Løkken: E-mail: kristin-hammarstrom.lokken@ffi.no

Automatic Target Recognition XXX, edited by Riad I. Hammoud, Timothy L. Overman, Abhijit Mahalanobis, Proc. of SPIE Vol. 11394, 113940W · © 2020 SPIE
CCC code: 0277-786X/20/$21 · doi: 10.1117/12.2558506

a vessel for more than a short period of time, the question becomes what to do in order to render detection, tracking or classification as hard as possible.

One approach that we will explore in this work is that of using adversarial camouflage (AC). Under this paradigm, the appearance of the vessel we wish to protect is structured in such a way that it confuses the software analyzing the images of the vessel.

Recent works in the domain of deep learning have shown that deep learning based algorithms for image analysis can be sensitive to surprisingly small changes in the images they analyze. Such techniques, typically described as adversarial techniques, have shown a considerable potential for fooling neural networks in a number of recent works.

In a previous work,[1] we have shown that a careful structuring of the visual appearance in grayscale images of a naval vessel can confuse deep learning based vessel classification algorithms to a substantial degree. In particular, we showed how even relatively small patches of very specific visual patterns, displayed on parts of a naval vessel, will render classification of that vessel much harder for a deep learning based vessel classifier.

In this work we look into how robust these patches actually are. How much can the resolution change and still confuse the classification net? How much can the contrast be reduced? What about image noise (including clutter)? The azimuth angle? In this paper such questions will be addressed.

In Section 2 we will give a short and very brief historical introduction to military vessel camouflage. We also provide an overview of the existing body of work related to techniques aimed at confusing neural networks, so called *adversarial* techniques. In Section 3 we will describe our neural network based approach to generating adversarial patterns for application to military vessels. We will also detail the robustness tests. In Section 4 we present and discuss the results we have obtained and in Section 5 we conclude.

2. RELATED WORKS

2.1 Naval vessel camouflage

In times of warfare, misleading the enemy, also called military deception, is of the essence. 'Deception' can be defined as the act of causing someone to accept as true or valid what is false or invalid.[2] Camouflage is regarded as a means to this end. At sea, camouflage can be divided into two categories:[3–5]

- Concealment or signature reduction – measures taken to blend in with the background.
- Disruptive type – artifices designed to deceive enemy sensor systems, rendering identification or targeting more difficult by making the size, range, speed, heading or class difficult to determine.

The value of camouflage as a means to thwart visual detection is eminently illustrated by the evolution of different species through natural selection. An overview of camouflage inspired from a zoological perspective can be found in Cuthill (2019).[6]

From a military perspective, different types of camouflage are cost-effective means of increasing survivability and combat persistence and have been used by armed forces all over the world throughout history.[3] Military naval vessels are hard to camouflage successfully. This is so both because of the sheer size of these agents, but also due to the variations in the warfare theaters where they operate – with light conditions, sea states and weather types ever changing.

In the naval warfare theater this has been a well known problem for a long time. During the First World War the British realized that the monochrome gray paint typically applied to their naval vessels would not effectively hide them from German submarines. In an attempt to remedy this situation, the Royal Navy introduced so-called dazzle camouflage, consisting of stripes, geometric patterns and eye-catching colours in different combinations. Experiments had shown that vessels with such designs were more difficult to classify as to type, speed, distance and bearing - both with the naked eye and through the optical distance gauges of German warships. Figure 1 shows an example of such dazzle camouflage.

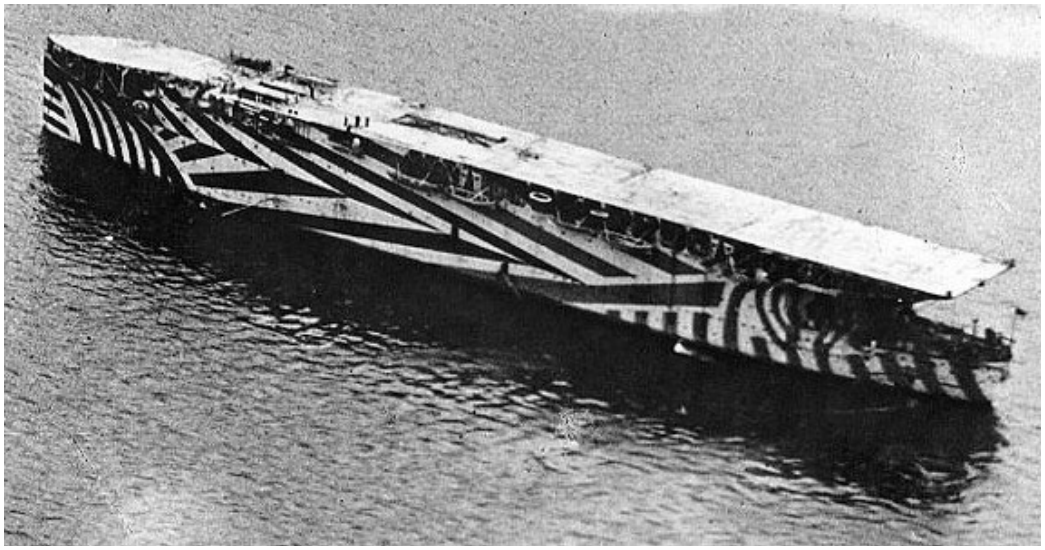


Figure 1: An example of dazzle camouflage. HMS Argus painted with dazzle camouflage in 1918. Photograph from wikipedia.org.[7]

Experiments with variants of dazzle camouflage were carried out also during the Second World War, see for instance López (2019).[8] The history of dazzle camouflage is described in more detail in USNI News.[4]

An interesting aspect with dazzle camouflage was that it did *not* necessarily aim at making the vessel harder to detect, but rather to make it harder for an observer to obtain good bearing, speed and distance estimates. As such it bears a certain resemblance to the type of camouflage we develop in the work reported here.

2.2 Neural networks

Recent developments have shown the value of different types of neural networks for a number of complex applications in image processing, see for instance Goodfellow (2016).[9] In particular, variants of the so called Convolutional Neural Networks (CNNs) have revolutionized the performance of automatic systems for object classification based on images. Today, performance of such systems is often on par with or even better than human performance (ibid). An interesting recent development in research related to neural networks is the introduction of adversarial attacks aimed to thwart the performance of a neural network.[10–16]

Recent work shows that small – undetectable to a human observer – changes in an image can be enough to fool a neural network discriminator, causing it to misclassify an image in a very dramatic fashion.[10–12,16,17] A recent study also produces images that are unrecognizable to human beings, but are classified as specific objects by state-of-the-art neural networks, with very high confidence.[18] This weakness in neural networks is shown to be due to the linear features of the neural networks.[12]

A further development of adversarial attacks is the concept of adversarial patches[14] and physical world attacks[13,19] on neural networks. It is shown that small patches with very specific patterns can make an image processing neural network misdetect or misclassify an object in the image. Such patches may fool the neural network even if they cover a relatively small part of the image, and regardless of whether or not they cover any of the features of the original object. For further reading about generative adversarial networks, see Goodfellow et al (2014b).[20]

In our previous work we used a generative adversarial network to create a certain pattern for the adversarial camouflage. Our approach consisted in developing specific patches that can be applied to naval vessels in order to make classification networks misclassify them. The question was whether it was possible to alter a military vessel (with paint for instance), in a way that would fool a neural network discriminator into misclassifying it as civilian. We demonstrated that it was possible, but we did not investigate the robustness of the patches. In

this work we demonstrate the impact that various changes in conditions – including resolution, contrast and orientation – have on the performance.

3. METHOD

In this section we will briefly describe our previous work[1] before we describe the robustness tests. During our work with the tests we developed a method to increase the robustness of the adversarial camouflage. This method is also presented here.

In order to design the patches, we use a neural network *generator* together with a neural network *discriminator*. Put together, this gives a generative adversarial network. A carefully designed patch from the generator can, when added to a limited smooth area on the hull and/or superstructure of a military vessel, fool a discriminator into misclassifying it as civilian. In the following, we will refer to such a patch of generated pattern as a patch of adversarial camouflage (AC).

3.1 Data

In order to train the networks involved in our experiments we need large numbers of images of maritime vessels, both civilian and military. One source for such images is the excellent web site `shipspotting.com`, which is a web site for shipspotters from all over the world. As a user one can upload ship images along with the location of the shot, the name and type of the vessel, etc. There are millions of images available and a large fraction of the images show military vessels, and the range of vessels is enormous.

We have downloaded a large number of images from shipspotting.com in the form of RGB images. These show vessels from all aspect angles as well as in all kinds of operating scenarios: in ports, in harbors, close to land, at sea, etc.

Prior to being fed to the neural networks, all images are rescaled to 800 rows by 1000 columns and converted to grayscale.

3.2 Discriminator

We have trained a convolutional neural network to function as the discriminator. The task of the discriminator is to analyze an input grayscale image of a maritime vessel and determine whether the vessel is civilian or military. The network is implemented in Keras.[21] It comprises five convolutional layers and two fully connected (dense) layers. The output of the last layer is a two-dimensional vector containing the probabilities that the vessel belongs to one of the two classes civilian or military.

The discriminator network was trained on a total of 120.000 images: 80.000 images of civilian vessels and 40.000 images of military vessels. For both classes, civilian and military, a total of 5 % of the images were reserved for validation at the end of each training epoch. At the end of training, the performance on the validation set is 95.5 % correct.

3.3 Generator

In order to train the generator, we select a total of 500 images of frigates (vessels clearly belonging to the class military). In each of these images we manually designated an area within which the generator may modify the images. Images of six frigates and the designated masks are shown in figure 2. Notice that the masks are slightly blurred (prior to use the borders of the masks are blurred by a Gaussian filter with parameter $\sigma = 5.0$). Also, notice that the masks have their opacity set to 70 %. Both these steps are taken in order to simulate a situation where the adversarial pattern is actually painted onto the vessel.

The aim of the generator is to produce a pattern that, when mixed into the frigate image within the area designated by the mask, will fool the discriminator into misclassifying the image of the frigate as a civilian vessel. Details of this process are given in,[1] where the discriminator was tricked into misclassifying images of military frigates (with embedded adversarial patches) as civilian ships.

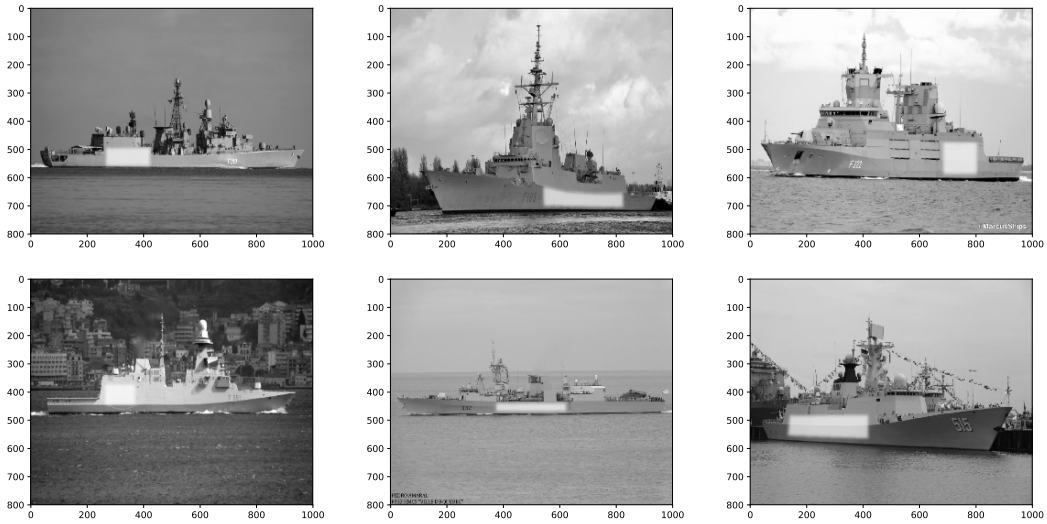


Figure 2: Images of six frigates with designated mask areas. All original photographs from shipspotting.com. Photographers (in reading order): *D173457Q Brian*, *Marcel and Ruud Coster*, *Marcus-S*, *Tomasello Letterio*, *Pedro Amaral* and *Ulf Kornfeld*.

3.4 Initial results

Our initial results are presented in.[1] These results indicate that such camouflage can be highly effective, as all of the six example frigate images were classified as civilian with a 100 % certainty, as shown in figure 3. Of the 500 images of frigates that were used to train the generator, 465 (93 %) were classified as civilian.

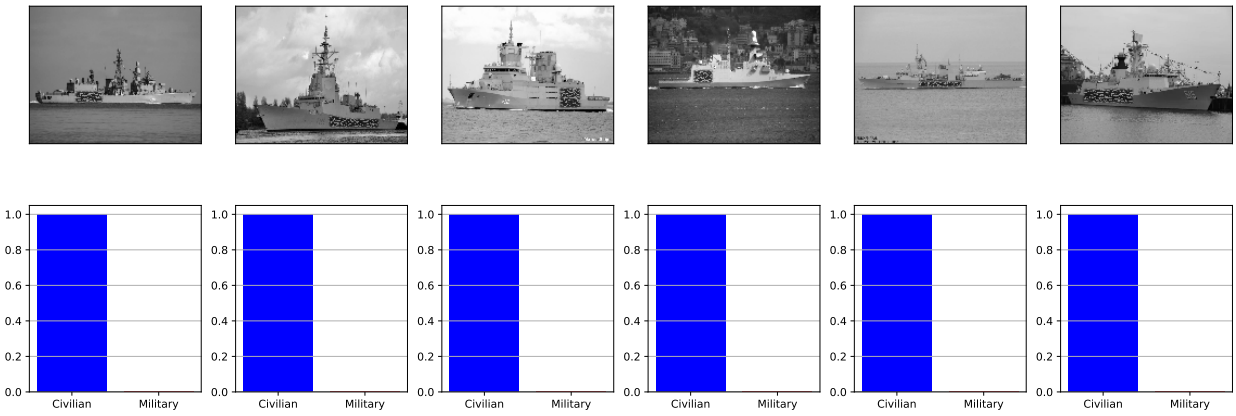


Figure 3: Images of six frigates with original patches inserted. All of the frigates are classified as civilian with 100 % certainty. All original photographs from shipspotting.com. Photographers (in reading order): *D173457Q Brian*, *Marcel and Ruud Coster*, *Marcus-S*, *Tomasello Letterio*, *Pedro Amaral* and *Ulf Kornfeld*.

It must be emphasized that our adversarial attack on the discriminator neural network is a so called white-box attack. In a white-box attack the attacker has some form of access to the parameters of the network that is to be fooled. In our case, the discriminator that is fooled by the adversarial camouflage is the same as the discriminator that is used for the training of the generator.

3.5 Robustness

Our aim in this article is to determine the robustness of the generated adversarial patches. How well will they perform when the ship is spotted at a distance or at a different angle? What if the image becomes degraded by

noise, clutter or low visibility? Would smaller patches yield the same result? In order to test this, we have carried out a number of experiments. The outputs from the generator have been degraded in various ways described below, and the following results from the discriminator monitored.

The images used in these tests are the same as the training set for the generator: 500 images where a mask has been hand drawn on the side of the ship, to insert the generated patches.

To illustrate the differences between the various degradations, we use one image of a frigate as an example. However, the tests have been applied to all 500 training images. The original output from the generator for the example image is shown in figure 4. Given that the patch fills a relatively small part of the image, we zoom in on the patch area to make the illustrations more readable. A zoomed-in image of the original output is shown in figure 5.

When testing the robustness, we have only made changes for the patches, i.e. the rest of the images remained unchanged. The reason is simply that we want to study the effects that various degradations have on the efficiency of the patches.

Figure 4: The original generator has created a patch for this image. Original photograph from shipspotting.com. Photographer: *Frank Schlünsen*

We have carried out the following degradation experiments. Examples of the various filters are shown in figure 6.

- Reduced resolution (a)
- Reduced contrast (b)
- Change in azimuth angle (c)
- Noise (d)
- Clutter (e)
- Erosion (f)

Figure 5: Zoomed in on the output from the original generator. Original photograph from shipspotting.com. Photographer: *Frank Schlünsen*

The aim of the *reduced resolution* is to simulate the patch at a longer distance. A longer distance means a smaller image, but as a fixed image size is needed as input to the neural net, an upsampling must be done. Thus, the procedure is first to reduce the image size, and then expand it. For both cases, a bicubic interpolation is applied. Finally, additive noise is included, so that the image noise in the modified patch is the same as the noise in the original patch.

Sometimes the visibility is poor. Thus, we want to investigate the robustness for such situations. The intensity value for a pixel in a patch with *reduced contrast* is calculated as

$$p_{rc} = \beta p_{org} + (1 - \beta)\hat{\mu} + N \tag{1}$$

where p_org and p_rc is the intensity for a pixel in the original patch and in the patch with reduced intensity respectively, $\hat{\mu}$ is mean intensity in the patch, $0 \leq \beta \leq 1$, and N is Gaussian noise. The noise is added so that image noise remains the same after the contrast has been reduced.

Another interesting situation occurs when there is a *change in azimuth angle*. How much can the azimuth change before the patch fails to fool the neural network? An α degree change in azimuth angle corresponds to resize the patch in the abscissa with a factor $\cos\alpha$. Figure 6 shows the result of simulating a change of 50°. The width of the patch is reduced and the pattern is squeezed, so as to mimic a 50 degree turn of the vessel.

We have added *noise* with standard deviation $\sigma = 20, 30, 40, 50, 60, 70, 80$ LSB. Since we're having 8 bit images, this means that the standard deviation is in the interval $8\,\% - 31\,\%$ of the maximum intensity.

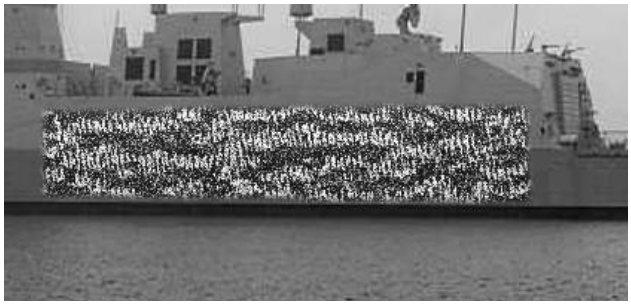
Clutter means in our context noise spots somewhat larger than pixel-to-pixel independent noise. These spots are multiplied (or added) to the patch. The following procedure generates clutter which qualitatively is found to look reasonable.

1. Generate a noise image consisting of Gaussian noise.
2. Reduce the image size with a factor 8.
3. Increase the image size with a factor 8 (bicubical interpolation used in these steps).
4. Multiply this clutter image with the patch image.
5. Clip the product so the intensities are in the interval $[0, 1]$.

(a) Reduced resolution: Resolution reduction factor 1.75.

(b) Reduced contrast: Contrast reduction factor 1.75.

(c) Azimuth angle: Mimicking a 50 degree turn of the vessel.



(d) Noise: Noise level 80.

(e) Clutter: Clutter level 90.

(f) Erosion: Height and width reduced by 50 percent.

Figure 6: Degradation of the output from the original generator. Original photograph from shipspotting.com. Photographer: *Frank Schlünsen*

Erosion is not exactly a degradation which can occur, but it is of interest to study change in performance as a function of the reduction of the patch size. We have determined the performance for 5, 10, 20, 25, 35, and 50 percent reduction. These reduction numbers refer to the width of the patch. The height is reduced equivalently during erosion.

3.6 Improved robustness

The patches we have generated so far, are outputs from the original generator only. A given patch is adapted to a particular image – a particular image of a particular vessel at a particular distance and orientation. An interesting question is whether the robustness towards various degradations can be increased.

We have trained new generators in order to produce more robust patches. This is done by adding various degradation filters at the output of the generators at training time. Apart from the newly introduced degradation filters, the training network – with generator and discriminator – are equal to the training of the original generator. The modified training network is illustrated in figure 7. The new filter is shown as a yellow block in the diagram.

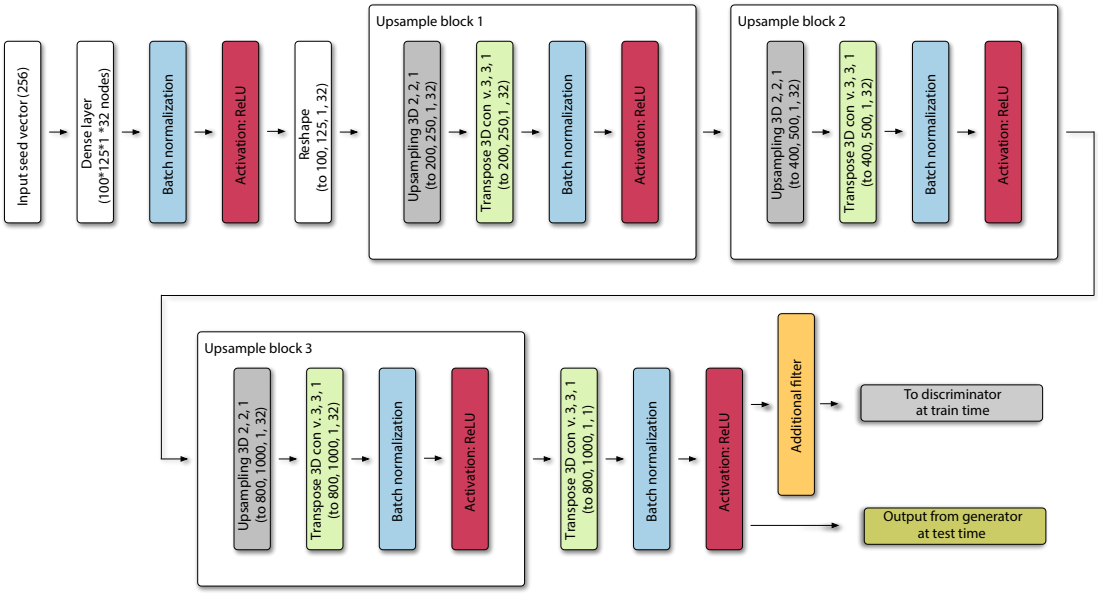


Figure 7: The original generator has been fitted with a degradation filter (yellow) on the output during training time. The filter is not trainable.

In the training of the original generator, the test output (green) was fed directly to the discriminator. In the current training, a filter (yellow) has been introduced between the output of the generator and the input of the discriminator. This filter modifies (i.e. degrades) the generated image before it is presented to the discriminator. This is done in order to force the generator to make patches that are more robust to this kind of degradation.

Due to limitations in Tensorflow, we have only been able to perform some preliminary tests. Two types of filters have been introduced: Image resolution reduction and image noise. The image resolution is reduced by resizing the images. That is, first they are made smaller, and then they are upscaled to the original size, using bilinear interpolation, both using the Tensorflow Keras Resize layer. The added noise is the Tensorflow Keras GaussianNoise layer.

4. RESULTS AND DISCUSSION

4.1 Results from testing outputs from the original generator

We first describe the results from testing the outputs from the original generator, before we started our work on improved robustness of the generator.

Having trained the complete adversarial network, we retrieve the "patch images" from the generator for each frigate image. These images, having the same size as the input image, are in turn modified as described in the previous section, before they are "ANDed" with the mask and overlaid the original image (70 % opacity used). The mask contour was blurred with a Gaussian filter ($\sigma = 5$, window size 7×7 pixels). The resulting images are similar to the regular outputs from the generator, except that the patch has been modified. These patch-modified images are fed to the discriminator, and the number of vessels classified as civilian ships is determined. The results from the experiments with degrading the patches are shown in figure 8.

The results from reduced resolution clearly show that even small changes reduce the patches' ability to fool the discriminator. A reduction of 1.1 gives that 10 % of the patches don't fool the discriminator. A reduction of 1.5 gives that half of the patches do not fool it, and if we have a reduction factor of 2, almost no patches fool the discriminator.

The results from the experiments with contrast reduction are pretty similar to what we observed for reduced resolution. With only small changes in contrast, many of the patches will not fool the discriminator.

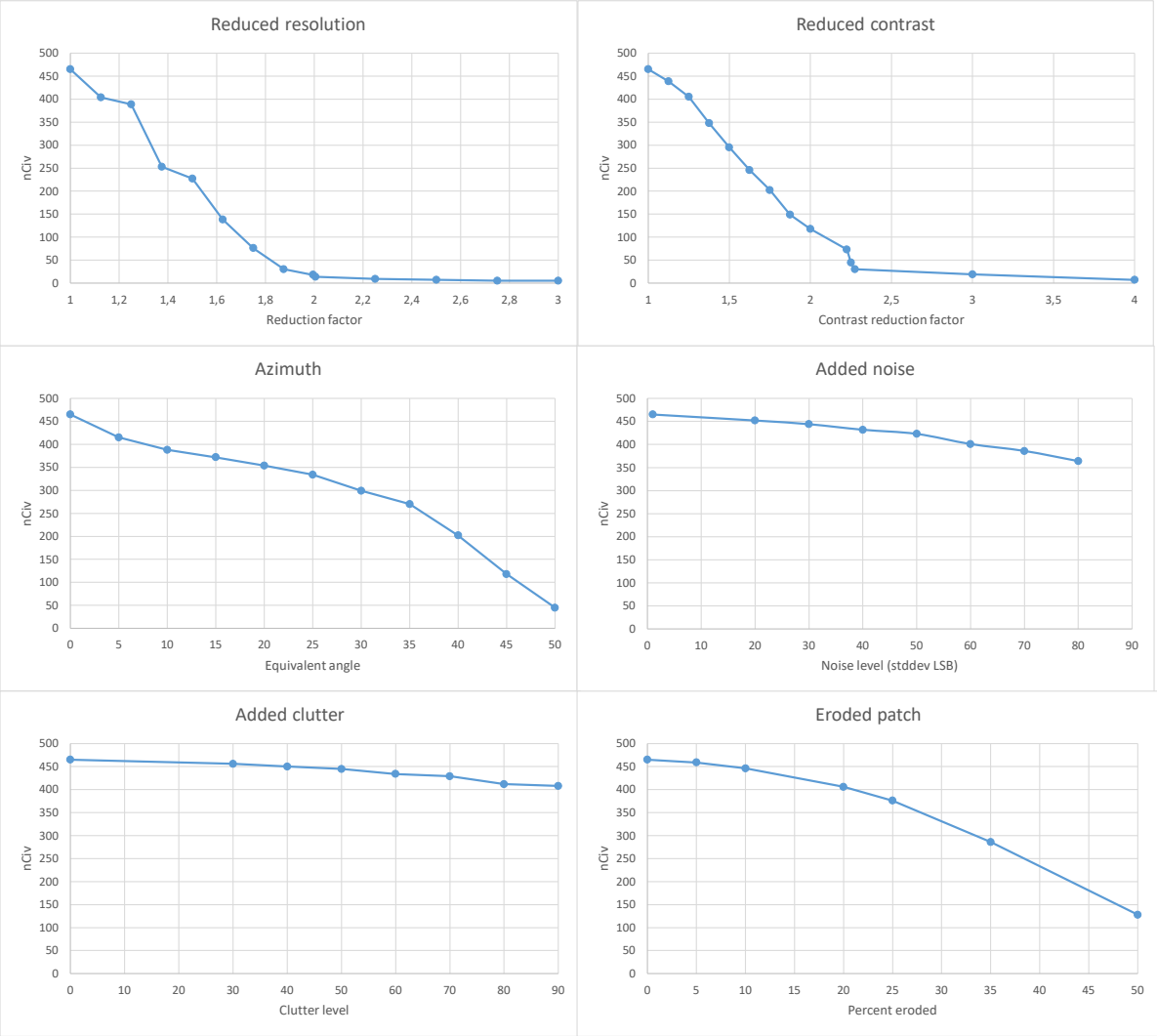


Figure 8: Images produced by the original generator have been degraded using various filters. The results from the discriminator are plotted here. On the x axes: Image degradation factors. On the y axes: Number of frigates classified as civilian.

From the figure is looks like the patches are less sensitive to changes in azimuth angles than for changes in resolution or contrast. However, one must remember that from a side view, a $25° - 30°$ change in azimuth is hardly visible from a low elevation angle, but it will result in a $28\,\% - 35\,\%$ reduction in number of patches being able to fool the discriminator.

The results from the noise experiments show that noise levels of $20 - 30$ LSB – i.e. a standard deviation of around $10\,\%$ of maximum intensity – has little impact on the results, even though this may be characterized as relatively high noise. Moreover, by increasing the noise levels to as much as 80 LSB – corresponding to $31\,\%$ of maximum intensity – we still see that the vast majority of the patches fool the discriminator.

The results from the clutter experiments show that the patches are resistant to clutter. The reason could be the same as for the noise, both the clutter as well as the noise are symmetrical.

The erosion experiments are also presented in figure 8. As long as $10\,\%$ or less of the width of the patch is

eroded, it seems that most of the patches are able to fool the discriminator. If 25 % or more of the patch width is eroded, less than 75 % of the patches will fool the discriminator.

4.2 Improving robustness of the generator

We have trained new generators, this time with degradation filters between the generator output and the discriminator input as shown in figure 7. These generators are trained in order to be more robust against degradation of the output images. These generators might be expected to produce patches that are more robust to degradation of the generated patches at test time. The resulting generators do indeed generate patches that are slightly different from the original generator. An example is shown in figure 9. The image shows the output from the original generator, the output from a generator trained with a resolution reduction filter on the output and the output from a generator trained with a noise filter of the output.

We have trained several new generators this way, with varying degrees of resolution reduction and noise levels on the output.

(a) Original generator.

(b) Generator trained with resolution reduction filter f_1.67.

(c) Generator trained with added noise filter f30.

Figure 9: Image from original generator (top left), image from a new generator trained with a resolution reduction filter between generator and discriminator (top right) and image from a new generator trained with added noise between generator and discriminator (bottom). Original photograph from shipspotting.com. Photographer: *Frank Schlünsen*

The new generators perform equally well on the six test images as the original generator. I.e. all six frigates are classified as civilian with 100 % certainty. The tests described in this section are performed on the 500 training images, as described above.

The output images from the resolution reduction trained generators have been tested against the previously described resolution degradation filters. The result is shown in figure 10. The blue line shows the performance of the original generator when the patch is degraded by the various resolution reduction filters. The plot shows that the generators trained with resolution reduction filters perform better than the original generator, in addition to being more robust against resolution reduction at test time. Resolution reduction with a factor 2 introduces an artefact which we have removed by using filters with factor 1.995 and 2.005 instead. Early plots showed high

spikes at factor 2, which are believed to be due to filter artefacts, and thus not relevant to our study. A similar effect might also cause the poor results of the f_2.0 generator training filter.

The output images from the noise trained generators are tested against the noise filters. The results are shown in figure 11. It is clear that the generators that are trained to withstand noise perform better when exposed to noise at test time.

We clearly see that the patches can be made more robust against small changes in various kind of degradations.

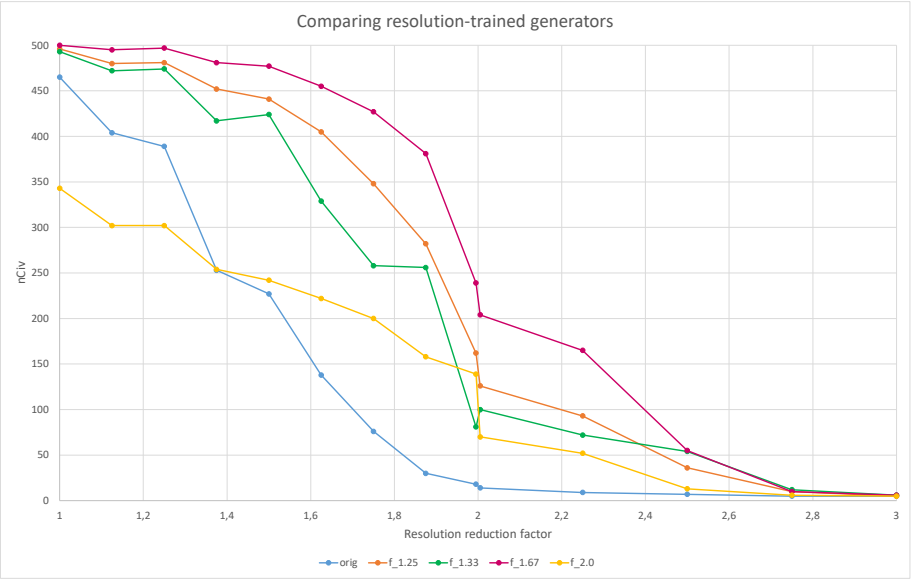


Figure 10: Comparing resolution-trained generators. Several generators have been trained, with varying resolution reduction filters between the generator output and the discriminator input. The plots show the robustness of these generators against resolution reduction.

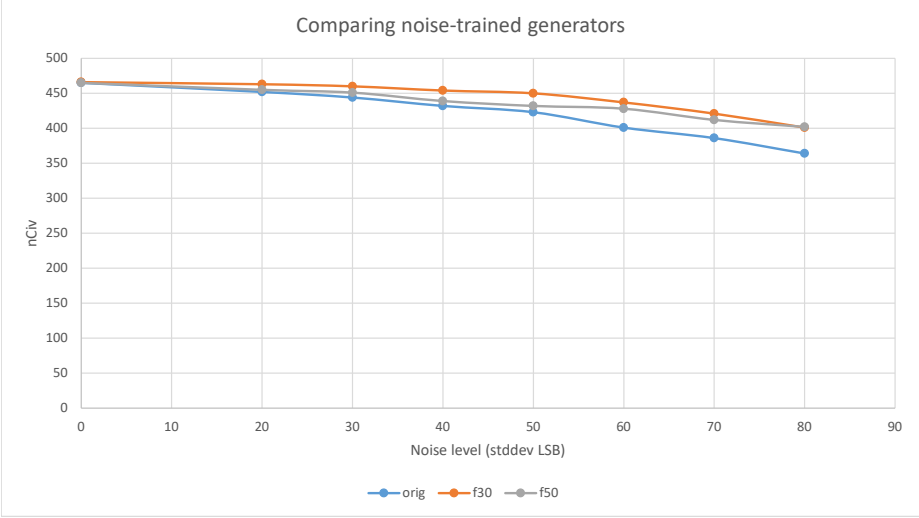


Figure 11: Comparing noise-trained generators. Several generators have been trained, with varying added noise between the generator output and the discriminator input. The plots show the robustness of these generators against added noise at test time.

So far, we have only been able to perform a few initial tests. But based on these tests, it's reasonable to be optimistic. We hope to report more results in near future.

5. CONCLUSIONS

In the work reported here we have investigated the use of generative neural networks in order to produce adversarial camouflage that will make discriminative neural networks trained to distinguish between civilian and military vessels fail. Specifically, we have investigated whether images of frigates, modified with patches of adversarial camouflage, will be misclassified as civilian vessels.

In this paper we have studied how robust such patches are against changes in features like resolution, contrast, noise, clutter, azimuth, and erosion. Results so far indicate that the patches are sensitive to many of these features. This means that small changes in e.g. resolution will have a large impact on the ability to fool the discriminator. The patches seem to be more robust against added noise or clutter than to degradations of contrast or resolution. An early understanding of this phenomenon is that both the clutter as well as the noise are symmetrical. We also observe that the black and white levels of the patches are intact for the noise and clutter images, while the reduced resolution and contrast patches have more gray level pixels, making them less prominent. Further work may be done to explore the robustness of the generator.

We have started work in order to find out whether a patch can be made more robust. A filter added between the generator and discriminator is applied. Initial results indicate that an increase in performance can be achieved by adding degradation filters between the generator and the discriminator at training time.

It should be pointed out that our work is a very early step in the assessment of AC as a tool for camouflage of military vessels. Our current results are based on a white-box attack, with only one military vessel class in the training set for the generator. A number of experiments must still be carried out in order to determine how general a tool AC actually is, and how suited it is for black-box attacks.

REFERENCES

[1] Aurdal, L., Løkken, K. H., Klausen, R. A., Brattli, A., and Palm, H. C., "Adversarial camouflage (ac) for naval vessels," in [*Proceedings of SPIE - The International Society for Optical Engineering*], **11169**, 111690H (09 2019).

[2] Merriam Webster Dictionary, "Deception." https://www.merriam-webster.com/dictionary/deception (2019). [Online: accessed August 8th 2019].

[3] Raven, A., "The Development of Naval Camouflage." http://www.shipcamouflage.com/development_of_naval_camouflage.htm (2019). [Online: accessed August 13th 2019].

[4] USNI News, "Camouflaged Ships: An Illustrated History." https://news.usni.org/2013/03/01/camouflaged-ships-an-illustrated-history (2013). [Online: accessed August 13th 2019].

[5] Merriam Webster Dictionary, "Camouflage." https://www.merriam-webster.com/dictionary/camouflage (2019). [Online: accessed August 8th 2019].

[6] Cuthill, I. C., "Camouflage," *Journal of Zoology, Thomas Henry Huxley Review* **308**, 75–92 (05 2019).

[7] Wikipedia, "Dazzle camouflage." https://en.wikipedia.org/wiki/Dazzle_camouflage (2019). [Online: accessed June 27th 2019].

[8] López, M., "Bismarck line drawings and paint schemes." https://www.kbismarck.com/drawings.html (2019). [Online: accessed August 13th 2019].

[9] Goodfellow, I., Bengio, Y., and Courville, A., [*Deep Learning*], The MIT Press, Cambridge, Massachusetts, USA (2016).

[10] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R., "Intriguing properties of neural networks," in [*International Conference on Learning Representations*], (2014).

[11] Carlini, N. and Wagner, D., "Towards evaluating the robustness of neural networks," in [*2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA*], 39–57 (2017).

[12] Goodfellow, I. J., Shlens, J., and Szegedy, C., "Explaining and Harnessing Adversarial Examples," *arXiv e-prints* , arXiv:1412.6572 (Dec 2014).

[13] Eykholt, K., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Xiao, C., Prakash, A., Kohno, T., and Song, D., "Robust Physical-World Attacks on Deep Learning Models," *arXiv e-prints* , arXiv:1707.08945 (Jul 2017).

[14] Brown, T. B., Mané, D., Roy, A., Abadi, M., and Gilmer, J., "Adversarial Patch," *arXiv e-prints* , arXiv:1712.09665 (Dec 2017).

[15] Sharif, M., Bhagavatula, S., Bauer, L., and Reiter, M. K., "Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition," in [*Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*], *CCS '16*, 1528–1540, ACM, New York, NY, USA (2016).

[16] Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Berkay Celik, Z., and Swami, A., "The Limitations of Deep Learning in Adversarial Settings," in [*2016 IEEE European Symposium on Security and Privacy (EuroS&P)*], 372–387 (2016).

[17] Gu, S. and Rigazio, L., "Towards Deep Neural Network Architectures Robust to Adversarial Examples," *arXiv e-prints* , arXiv:1412.5068 (Dec 2014).

[18] Nguyen, A., Yosinski, J., and Clune, J., "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images," in [*Computer Vision and Pattern recognition (CVPR '15)*], IEEE (2015).

[19] Thys, S., Van Ranst, W., and Goedemé, T., "Fooling automated surveillance cameras: adversarial patches to attack person detection," *arXiv e-prints* , arXiv:1904.08653 (Apr 2019).

[20] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative adversarial nets," in [*Advances in Neural Information Processing Systems 27*], Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., eds., 2672–2680, Curran Associates, Inc. (2014).

[21] https://keras.io/.