

# Interpolation merge as augmentation technique in the problem of ship classification

Dawid Połap\* and Marta Włodarczyk-Sielicka†

\*Marine Technology Ltd.

ul. Roszczyńskiego 4/6, 81-521 Gdynia, Poland

†Maritime University of Szczecin

Waly Chrobrego 1-2, 70-500 Szczecin, Poland

Email: \*d.polap@marinetechonology.pl, †m.wlodarczyk@am.szczecin.pl

**Abstract**—Quite a common problem during training the classifier is a small number of samples in the training database, which can significantly affect the obtained results. To increase them, data augmentation can be used, which generates new samples based on existing ones, most often using simple transformations. In this paper, we propose a new approach to generate such samples using image processing techniques and discrete interpolation method. The described technique creates a new image sample using at least two others in the same class. To verify the proposed approach, we performed tests using different architectures of convolution neural networks for the ship classification problem.

## I. INTRODUCTION

IMAGE classification is considered one of the leading problems in today's AI where many inconvenient situations may occur. One such element is too much variety of objects belonging to the same class. Besides, each of these objects can be placed on a different background, in a different light the same as from a certain angle. Another one is too small several samples for selected classes to train the classifier at a given level of accuracy. Unfortunately, there is also the situation of overtraining the model which means correct classification at the training database level, but the large probability of incorrect classification for new samples. These are only selected problems encountered in the classification task.

To reduce the likelihood of any of them occurring, a proper detection [1] or data augmentation can be used. It is a process of increasing the number of samples in the database using existing ones. This kind of generating process is mainly based on rotation, changing brightness or zooming. All of them are basic image processing techniques but provides more samples. It was proved that accuracy can be higher using this type of approach what can be seen in [2], where the authors analyzed the impact of augmented data on the effectiveness of convolutional neural network (CNN). Again in [3], the idea of generating data during the process of driving and using them to generate newer is described. A big problem in a small amount of data is medical data, where the diseases can be of different form and shape. In [4], the idea of using augmented cytological images of cells was presented. Similar ideas of using generating sample techniques with CNN were used in the target recognition problem in [5], [6]. In [7], the idea of using CNN for semantic segmentation was presented. In this solution, to increase the accuracy level of CNN,

augmentation was added to achieve wanted results. Moreover, in [8], authors used the occlusion technique, according to which random rectangles composed of pixels in the HSL color model are randomly obscured. This type of action also changing image, so as a result a new sample is generated. Another approach of augmentation solution was presented in [9], wherein the architecture of CNN, the authors added a layer that is responsible for this process. Similar idea was described in [10],

In this paper, we propose a novel technique of data augmentation for the classification process. Our approach is based on interpolated data from two images for merging into the third one as a hybrid one. This approach was tested in the problem of ship classification.

## II. INTERPOLATION MERGE TECHNIQUE

Having two images from the same class  $I_1$  and  $I_2$ , there is a way to merge them into a new image. At first, each of these images must be processed to calculate the merging ability. For this purpose, some image processing technique is used for simplicity and adaptation to interpolation.

Each image is subjected to convolutional operation which is a modification of image using some filter  $g$ . A filter is described as a matrix of size  $k \times k$ , and this operation can be described as following

$$I'[i, j] = I[i, j] * g = \sum_{t=-\lfloor k/2 \rfloor}^{\lfloor k/2 \rfloor} \sum_{r=-\lfloor k/2 \rfloor}^{\lfloor k/2 \rfloor} I[i+t, j+r] \cdot g[t, r]. \quad (1)$$

The coefficient of matrix  $g$  depends on the filter. In our proposition, each image is converted as

$$I'[i, j] = ((I[i, j] * g_1)^5 * g_2) * g_3, \quad (2)$$

where the filters  $g_i$  define Gaussian blur, emboss and edge detection as follows

$$g_1 = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \quad (3)$$

$$g_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}, \quad (4)$$

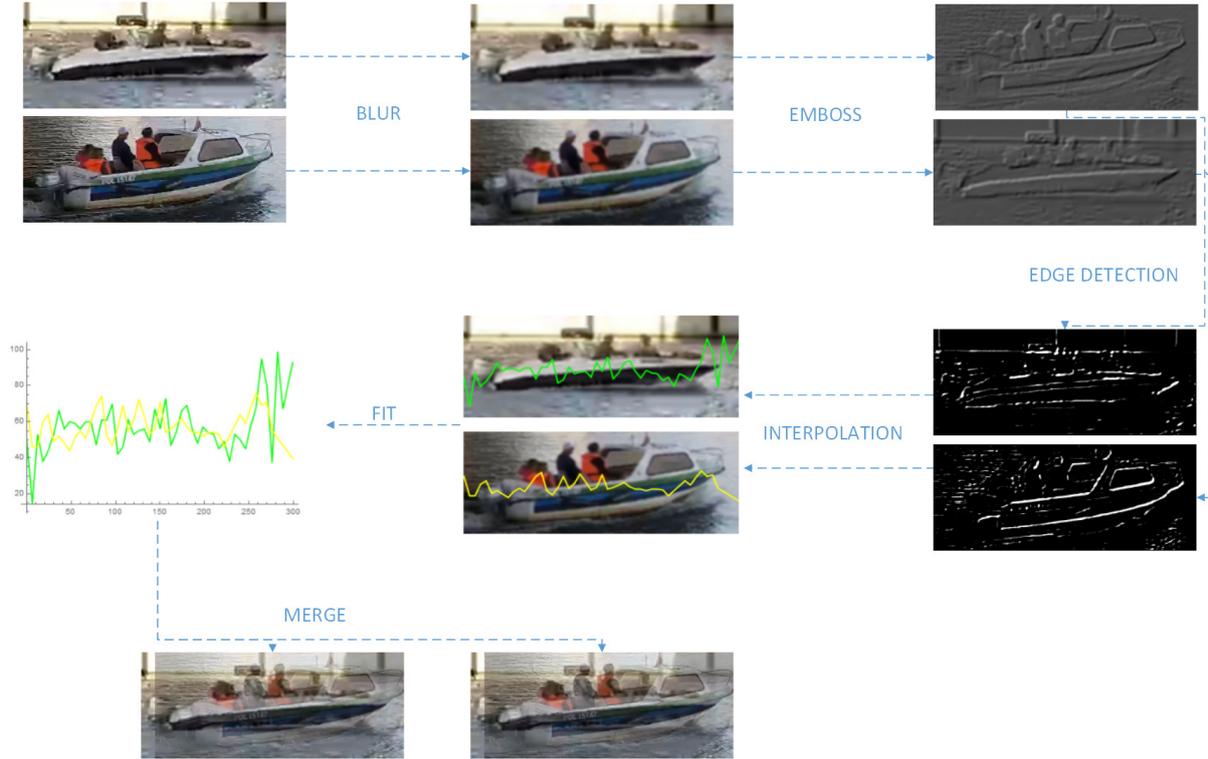


Figure 1: Visualization of the proposed augmentation technique.

$$g_3 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}. \quad (5)$$

The above filters are needed to obtain only some pixels that represent the shape of a ship. In Eq. 2, there is five times blur due to the removal of as much noise as possible and smaller elements, and above all water. The remaining not black pixels in obtained image  $I'$  must be corrected before using them in the interpolation technique which requires uniquely defined points. To achieve this uniqueness to the OX axis, for each column of pixel in  $I'$ , the average value of all values non-black pixels will be calculated according to

$$\left( \frac{1}{p} \sum_{i=1}^p x_i, y \right), \quad (6)$$

where  $p$  is the number of all non-black pixels in  $x$  column.

Obtained points can be defined as a set  $\Theta = \{(x_1, y_1), (x_2, y_2), \dots, (x_p, y_{p-1})\}$  where each point can be presented as  $y = f(x)$ . These points are interpolated by some function  $F(x) = \sum_{o=0}^m a_o \phi_o$  ( $\phi_o$  are a polynomial system). As interpolation, we understood finding all coefficient  $a_m$  of  $F(x)$  in such a way that error is the smallest one, what can be defined as

$$\|f - F\| = \sum_{i=1}^p w(x_i) [f(x_i) - F(x_i)]^2. \quad (7)$$

As a result, we obtain an interpolated functions  $F_{I_1}$  and  $F_{I_2}$ . Both functions are compared with each other to minimize the area between curves as

$$\int_0^{width} |F_{I_1}(x) - F_{I_2}(x)| dx. \quad (8)$$

When the error value is the smallest, both images are superimposed to fit the function offset. The described operation is performed concerning the width, although it can be done in the same way in height – then the fit will be more accurate. Image joining consists of changing the alpha channel  $\alpha$  on the shifted image, which will cause it to shine through. An example of augmentation is illustrated in Fig. 1.

Of course, it is possible to use basic augmentation techniques like zooming, rotation, etc, which will give more images with a more accurate fit.

### III. EXPERIMENTS

Described technique was implemented in Wolfram Mathematica 11 and C# language. For testing purposes, two classes representing cargo (2120 images) and military (1167 images) vessels from the Deep Learning Hackathon publicly available collection organized by Analytics Vidhya were used, and one class was created as part of the SHREC project [11] consisting of 688 images of yachts. This database was used for analyzing proposed technique for classical architecture of convolutional neural network like AlexNet [12], VGG16 [13] and Inception [14].

	CNN	10 iterations	15 iterations
without augmentation	AlexNet	0,62	0,74
	VGG16	0,64	0,75
	Inception	0,63	0,77
$\alpha = 0.2$	AlexNet	0,67	0,78
	VGG16	0,68	0,8
	Inception	0,59	0,82
$\alpha = 0.4$	AlexNet	0,64	0,85
	VGG16	0,62	0,86
	Inception	0,67	0,87
$\alpha = 0.6$	AlexNet	0,65	0,84
	VGG16	0,67	0,85
	Inception	0,71	0,87
$\alpha = 0.8$	AlexNet	0,59	0,8
	VGG16	0,61	0,76
	Inception	0,6	0,83

Table I: Comparison of the effectiveness of various CNN architectures in relation to the proposed augmentation and the number of training iterations (without freezing method).

Statistical coefficient	Value
accuracy	0,865454545
sensitivity	0,962760131
specificity	0,390374332
precision	0,885196375
negative predictive value	0,682242991
miss rate	0,037239869
fall-out	0,609625668
false discovery rate	0,114803625
false omission rate	0,317757009
F1 score	0,922350472

Table II: Statistical values for Inception architecture trained on the basis of data augmentation.

For each of class, 75% of samples were used for training purpose (with a ratio 80:20 – training:validation samples), the remaining 25% was used for verification purpose to analyze accuracy. Then, for each of class, 20% of images was chosen for augmentation, which allowed for a 10% increase in each class.

The images used for augmentation were selected randomly and created according to the described technique. As part of the analysis, the following degrees  $t$  of polynomials were used – 2, 3 and 4. Sample interpolation function with  $t$ -th degree are presented in Fig. 2. We notice, that for degrees higher than 2 and small samples, the function can reach extremum above or even below image which is not advisable to use – due to the later fit. When a grade 2 polynomial is used, the graph is more rounded, so matching between two graphs can be simpler but formally there are more calculations to find this polynomial.

Samples were created for second degree of interpolated polynomial. All samples has transparency coefficient  $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$ . Each of CNN was trained with this database for 10 and 15 iterations and then checked accuracy using the verification database (with additional 107 images of other types of ship).

The obtained results are presented in Tab. I. Based on these results, using proposed augmentation increases accuracy by an average of 7%. It is easy to see that the increase in the number of iterations resulted in an increase in accuracy for each architecture without as well as with augmentation. The



(a)  $t = 1$



(b)  $t = 2$



(c)  $t = 3$



(d)  $t = 4$

Figure 2: Samples with specific degree  $t$  of interpolation function.

best results were obtained for the transparency parameter value of 0.4 and 0.6. The reason for these results is the fact that at these levels both ships are visible with a minimum clearance relative to the other and are stacked on top of each other (by width). Also, for Inception architecture and transparency values of 0.4 as well as 0.6, the highest accuracy was reached at 0.87. Other statistical coefficients were calculated for this configuration and are presented in Tab. II and in Fig. 3. Despite the high accuracy and precision, the architecture showed that the sensitivity of the model is very high, as it is 0.96, and is understood as the probability that the classification will be correct if the entering image was correct. Also, the specificity was almost 0.4, which is not the best result, as it is the probability of correct classification for a false sample. Besides, the false discovery rate has reached a probability level of 0.11, which means that only 11% of the positive results will not be correct.

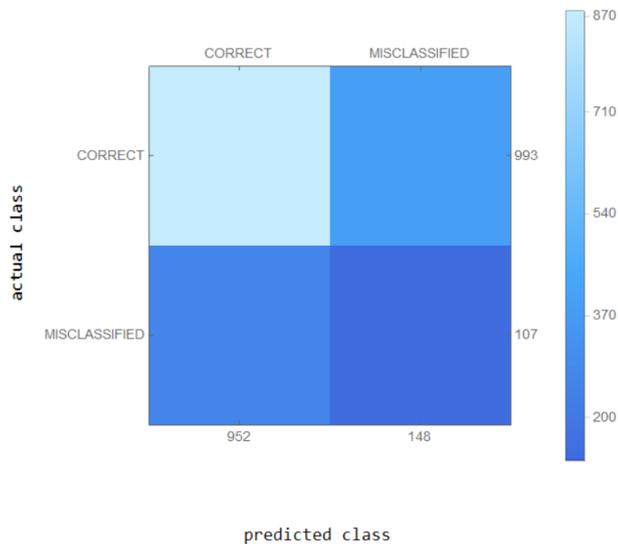


Figure 3: Confusion matrix for the tested Inception architecture.

The results indicate the correct impact of the newly generated data on the quality of the classification when performing tests on three classic architectures using learning transfer. Besides, the calculated statistical coefficients confirm that the convolution network model is properly trained and has high efficiency in detecting correct samples as well as a high probability of classification into a given class.

#### IV. CONCLUSION

Data augmentation is important in the training process when there is no sufficient amount of samples. In this paper, we propose an alternative approach for creating image data based on two other samples using merging images and locating them based on the interpolation technique. We showed that this approach can be useful for training a convolutional neural networks. Based on the obtained results, our approach reached an average accuracy value 7% better than the classic approach without using an extended database.

In the future, we plan to analyze other techniques of mixing two images which might prove to be a better way to generating new image samples.

#### ACKNOWLEDGMENTS

This work was supported by the National Centre for Research and Development (NCBR) of Poland under grant no. LIDER/17/0098/L-8/16/NCBR/2017.

#### REFERENCES

- [1] T. Hyla and N. Wawrzyniak, "Ships detection on inland waters using video surveillance system," in *IFIP International Conference on Computer Information Systems and Industrial Management*. Springer, 2019, pp. 39–49.
- [2] M. A. Kutlugün, Y. Sirin, and M. Karakaya, "The effects of augmented training dataset on performance of convolutional neural networks in face recognition system," in *2019 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2019, pp. 929–932.
- [3] W. Zhang, P. M. Chu, K. Huang, and K. Cho, "Driving data generation using affinity propagation, data augmentation, and convolutional neural network in communication system," *International Journal of Communication Systems*, p. e3982, 2019.
- [4] A. Teramoto, A. Yamada, Y. Kiriya, T. Tsukamoto, K. Yan, L. Zhang, K. Imaizumi, K. Saito, and H. Fujita, "Automated classification of benign and malignant cells from lung cytological images using deep convolutional neural network," *Informatics in Medicine Unlocked*, vol. 16, p. 100205, 2019.
- [5] N. J. Tustison, B. B. Avants, Z. Lin, X. Feng, N. Cullen, J. F. Mata, L. Flors, J. C. Gee, T. A. Altes, J. P. Mugler III *et al.*, "Convolutional neural networks with template-based data augmentation for functional lung image quantification," *Academic radiology*, vol. 26, no. 3, pp. 412–423, 2019.
- [6] F. Gao, T. Huang, J. Sun, J. Wang, A. Hussain, and E. Yang, "A new algorithm for sar image target recognition based on an improved deep convolutional neural network," *Cognitive Computation*, vol. 11, no. 6, pp. 809–824, 2019.
- [7] K. Cho *et al.*, "Retrieval-augmented convolutional neural networks against adversarial examples," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 563–11 571.
- [8] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and J. Li, "Hyperspectral image classification using random occlusion data augmentation," *IEEE Geoscience and Remote Sensing Letters*, 2019.
- [9] G. Chen, C. Li, W. Wei, W. Jing, M. Woźniak, T. Blažauskas, and R. Damaševičius, "Fully convolutional neural network with augmented atrous spatial pyramid pool and fully connected fusion path for high resolution remote sensing image segmentation," *Applied Sciences*, vol. 9, no. 9, p. 1816, 2019.
- [10] L. Mou, Y. Hua, and X. X. Zhu, "A relation-augmented fully convolutional network for semantic segmentation in aerial scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 12 416–12 425.
- [11] N. Wawrzyniak and A. Stateczny, "Automatic watercraft recognition and identification on water areas covered by video monitoring as extension for sea and river traffic supervision systems," *Polish Maritime Research*, vol. 25, no. s1, pp. 5–13, 2018.
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," pp. 1097–1105, 2012.
- [13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.