Preprocessing Method Comparisons For VGG16 Fast-RCNN Pistol Detection

Jiahao Li[1], Charles Ablan[1], Rui Wu[2], Shanyue Guan[1], and Jason Yao[1]

[1] Department of Engineering, East Carolina University, Greenville, North Carolina, U.S.A
lij17@students.ecu.edu
ablanc17@students.ecu.edu
guans18@ecu.edu
yaoj@ecu.edu

[2] Department of Computer Science, East Carolina University, Greenville, North Carolina, U.S.A
wur18@ecu.edu

Abstract

In recent years, gun detection and threat surveillance became a popular issue as gun violence continued to threaten public safety. Convolution Neural Networks (CNN) has achieved impressive gun detection precision with the advancements in graphic processing units. While many articles have proposed beneficial complex architectures within the neural network, there has been little study on effective image preprocessing techniques that supplement neural networks. With the objective of increasing neural net precision using image processing techniques, this research analyzes three different approaches to image preprocessing using a VGG16 trained Fast Regional Convolutional Neural Network (F-RCNN) pistol detector. The base VGG16 was trained with transfer learning in MATLAB on a dataset of 1500 pistol images and tested on 500 more. The results of the original VGG16 detector are compared with the results of the other VGG16 detectors trained with various image processing techniques to determine the viability of each technique. The three image processing techniques are as follows, color contrast enhancement, principle component analysis (PCA), and a combined preprocessing method. After testing the detector trained with the three methods above, it was found that the color enhancement technique had the best success in raising precision with proper levels of color contrast adjustments. The PCA analysis proved to be incompatible for the neural net to learn features of images that has not underwent PCA processing and thus the method failed to produce beneficial results on the unmodified testing dataset. The combined method processing took both PCA and color contrast enhancement techniques and combined the results into a single training dataset. The combined preprocessing method proved to be ineffective in raising precision potentially due to conflicting features.

1 Introduction

Past breakthroughs in Convolutional Neural Networks (CNN) using VGG16 and VGG19 architectures achieved an incredible 90 percent accuracy in image classification [21][3]. As the search for better neural networks continue, increasing complexity becomes unavoidable[3]. A

CNN's classification accuracy often increases with layer complexity but with it the speed decreases along with increased memory usage[6]. The new 1000+ layered Inception V networks are evidence that CNN is getting increasingly complex at the expense of memory and speed[24]. Maintaining the simplicity of the network architecture while achieving high identification precision becomes a growing concern in recent years [22]. Of the existing neural networks, VGG16 remained one of the simplest yet robust neural network ever created. This study applied a Fast-Regional Convolutional Neural Network (F-RCNN) model that modified a VGG16 net into an F-RCNN object detector for pistol detection[7]. The open source VGG16 net was trained using transfer learning within MATLAB to achieve an overall true positive percentage of over 90 percent on pistol detection. The focus of this study is to determine viable image preprocessing techniques that can increase detector performances without modifying the network's layers.

This experiment utilizes MATLAB to both train the F-RCNN object detector and implement various image preprocessing techniques. MATLAB was used because the Deep Learning toolbox is easy to implement with robust supports for various neural networks [23][9]. The built-in conversion function of images to matrix form made image processing easy and efficient as various image processing techniques involves matrix transform and matrix transpose[16][14]. Three image processing trials were tested on the training dataset for F-RCNN pistol detection in hopes of raising detection precision. A neural network performs the best if trained with great variance in pose and lighting of an object[27]. The image preprocessing techniques applied in this study operated under the assumption that pose and lighting are key features that greatly effects the precision of any object recognition systems [27]. Indeed, neural network's tendency to rely on color and lighting makes distinguishing dark simple objects from low lighting backgrounds challenging[17]. To resolve the problem of low color contrast with the background, the first trial used color enhancement techniques at set intervals to widen the gap between dark and light-colored objects. The second trial used PCA analysis to reduce the background while retaining most of the pistol's features[20][8]. After the results of the first two methods were obtained, a fusion of the two modified image datasets along with the original dataset was used to train a combined F-RCNN VGG16 detector.

In the rest of this paper, Section 2 describes the related works in the area of pistol detection and image preprocessing. Section 3 describes the dataset used in the study as well as the source images used. Section 4 describes the performance of the VGG16 F-RCNN detector on the base dataset. Section 5 describes the effects of various color enhancement trials on detector precision. Section 6 describes the effect of PCA analysis on detector performance. Section 7 describes the effect of combining the different image preprocessing techniques into one dataset. Section 8 discusses the results of the previous experimental trials. The conclusion is given in Section 9; and finally, the acknowledgments are given in Section 10.

2   Related Works

Pistol detection has been a widely researched topic ever since Neural Networks first gained popularity with Alex-net in 2012 [10].Many studies in the past has tackled the problem of handgun detection using Regional Proposal Networks and transfer learning from established neural networks such as VGG16 [2][17]. Two methods are generally used to improve the performance of specific object detectors. The first method is modifying the underlying neural network to become deeper and more robust [24], and the second method is applying image preprocessing techniques to help detectors to better separate the important features from background noise [19]. This paper tackles both the problem of creating an acceptable FAST-RCNN pistol detection network and using image preprocessing techniques to increase the network performance.

40

2.1   Pistol Detection

Pistol detection differs from pistol classification in that object detectors must identify the part of the image with the weapon[17][11][12][5]. Various pistol detectors has been trained in the past using both neural networks and traditional machine learning techniques such as SVM(Support Vector Machine) and Histogram of Oriented Gradients (HOG)[17][25]. However, the best performances came from using Regional Proposal Networks(RCNN) and its variants the FAST-RCNN and Faster-RCNN[17][5].

Akcay et al.[2] used a variety of neural networks with different types of object detectors to detect pistols in x-ray images. Five different object detection models were used in the study including Sliding Windows Convolutional Neural Network(SW-CNN), Regional Convolutional Neural Network (RCNN), Faster Regional Convolutional Neural Network (Faster-RCNN), Region-based Fully Convolutional Networks (R-FCN) and You Only Look Once object detectors (YOLOv2). The object detection models were trained with Alexnet, VGG16, VGG19 and residual neural networks (ResNets). The resultant data shows that both RCNN and Faster-RCNN outperformed both traditional handcrafted Bag of visual words(BoVW) features and fellow SW-CNN detectors. Akcay et al.[2] has proved with their study that it is possible to achieve high object detection precision with R-CNN models and their variants.

Olmos,Tabik and Herra[17] analysed a VGG16 based Faster-RCNN detector for video pistol detection with limited success. The Faster-RCNN's large false positives rates drastically lowered the overall detector precision. The researchers contributed the high false positives ratio to low contrast and luminosity of certain video frames. The pistols that are not clearly distinguished from the backgrounds are often missed and other objects are falsely identified.

2.2   Neural Network Image Preprocessing

Within the field of image analysis, image preprocessing techniques are frequently used in combination with a variety of image classification algorithms [26][28][19][5]. Many image preprocessing techniques aimed to either reduce noise or enhance desired features. Rehman, Amjad and Saba, Tanzila[19] discussed a variety of image preprocessing methods for character recognition using neural networks. Within the various preprocessing techniques discussed, the researchers highlighted the importance of thresholding in image processing. Thresholding sets a boundary on the original color scheme from which the image is converted to binary or gray scale. Threshold simplifies the image and highlights the desired characters making it an important preliminary image processing technique. The other image preprocessing techniques included eliminations of unwanted features and extraction of key features. The importance of image preprocessing techniques was highlighted by Rehman, Amjad and Saba, Tanzila[19] and its effect in increasing neural network performances cannot be overlooked.

3   Dataset

This study utilized 2000 open source images from various image datasets under an academic license as well as original images taken by the research team. Out of a training dataset of 2000 images, 1500 were randomly selected for training and 500 were used for testing. The first dataset used was the IMFDb online firearm dataset, a free gun image dataset that contains a variety of pistols, rifles, shotguns, and other firearms[15]. The second dataset used was the Sci2s weapons dataset from the university University of Granada[17][5]. For this experiment, only pistols were selected from a variety of movie images. The selection criteria for the combined

image dataset included a variety of background and contextual information that is crucial in training a robust neural network[18]. Finally, the image dataset is completed by 150 images taken by the research team. The research team took the images with 3D-printed pistols at various backgrounds, angles, and distances for a balanced selection of various pistol images. The combined image dataset was biased towards images with a variety of background and contextual information for accurate object recognition[18].

4   Base VGG16 F-RCNN Performance

Within MATLAB, a pre-trained VGG16 net was trained with the 1500 pistol images. The VGG16 object detector reached a training accuracy of 98.8 percent after 5 epochs. The trained detector was then tested on the 500 testing images with an average precision of 0.2138. The low precision likely resulted from boxing errors where either a part of the gun or too much of the background was selected. However, upon physical examination, it was found that although the detection boxes did not precisely match the ground truth labeling, over 90 percent of the highest confidence detection box correctly identified most of the weapon and did not mistake other objects for

Figure 1: VGG16 detection precision on 500 images

pistols. Thus, the overall accuracy of the base F-RCNN object detector was found to be in an acceptable range. Figure 1 shows the MATLAB's precision evaluation of the VGG16 during testing. Figure 1 shows the performance graph plotted for precision versus recall, where recall and precision are ratios of true positive instances to the sum of true positives and false negatives in the detector, based on the ground truth[17][2]. By running through each image and visually judging the accuracy of the neural net, three different categories metrics were used to determine true the performance of a neural net performance.: i) True Positive means that the network has correctly identified the weapon with an appropriate bounding box. ; ii) False Positive means that the network has failed to select most of the weapon or has mistaken something else for the pistol. ; iii) Negative means the Neural net has failed to recognize guns in the image with a high confidence. The results of the visual analysis will change slightly from person to person and trial to trial, but five successive repetitions on the testing dataset proves the general effectiveness of the trained F-RCNN object detector. Table 1 shows the classification criteria for accuracy evaluation and the average performance of the VGG16 detector over the five repetitions.
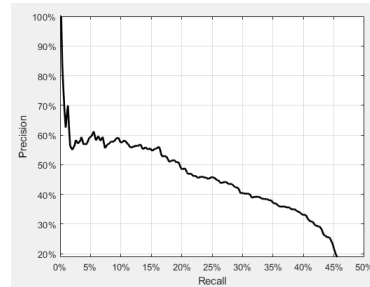
Table 1: VGG16 judgment criteria and percentage number

True Positive	False Positive	Negative
453/500	47/500	0/500

5   Color Enhancement Trial

The color enhancement preprocessing techniques operates under the assumption that neural nets relies heavily on color and texture for target identification[17][1]. Since most pistols have a darker color as opposed to their surroundings, by increasing the color contrasts the neural net should be able to better distinct dark pistols from light backgrounds to avoid false positives. A flowchart of the color enhancement process is shown in Figure 2.

The first preprocessing method separated light and dark colors at 75 out of 255 for all three RGB color schemes. The 75 separation value is selected because it separates the darker gun color scheme from the lighter background.The parts of the image with a RGB value less than 75 are darkened by -30 and the parts with a number larger than 75 are lightened by +30. The 30 enhancement value was chosen because initial trials indicated that any value less than 30 produced had no significant impact on the testing precision. If a particular pixel has a RGB value of less than 30 it will simplify to zero after preprocessing and the same concept applies to values greater than 225. The images with a mean RGB value of 85 or below were deemed too dark for the pistol to be separated from the background and thus the color enhancement was not performed.

Figure 2: Color enhancement flowchart.

Only the training dataset underwent this preprocessing technique and the testing results on the original testing dataset are shown in Figure 3. In the rest of this paper, the "x" of "enhanced-x" refers to the specific RGB value that is modified in the image.

From Figure 3 it can be observed that the enhanced-30 dataset did not result in significant improvements in MATLAB recorded average precision from 0.2138 to 0.2208. Since initial trials with the color enhanced-30 showed only slight improvements in precision, the second trial doubled the modification ratio to +60 and -60. From Figure 3 it is observed that the enhanced-60 trial noticeably increased the average precision from 0.2138 to 0.2944 for a 38 percent increase in precision. Based on the success of the enhanced-60 trial, the enhanced-90 trial further increased the contrast to +90 and -90 in attempt to achieve higher precision. As it can be observed from Figure 3, overly increasing the color enhancement ratio to +90 and -90 negatively impacted detector performance decreasing the average precision to only 0.0759.

Table 2 shows the full comparison of the effect of different color enhancement ratios on training images. It can be seen from the color enhanced images that the pistol has a drastically higher color contrast as opposed to the background. By using color enhancement techniques, the background noise is also whitewashed resulting in an overall simpler image.

6   PCA Feature Reduction

Each layer of a Convolutional Neural Network retains a specific feature of the image[6][24]. Because of the curse of dimensionality, some research results show that fewer features can be less misleading for a machine learning model. PCA feature reduction aimed to reduce the amount of
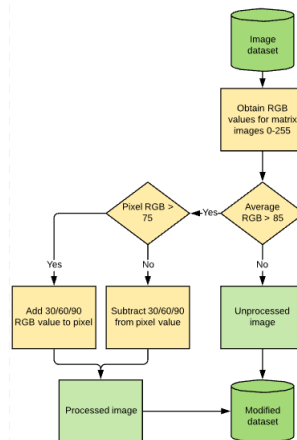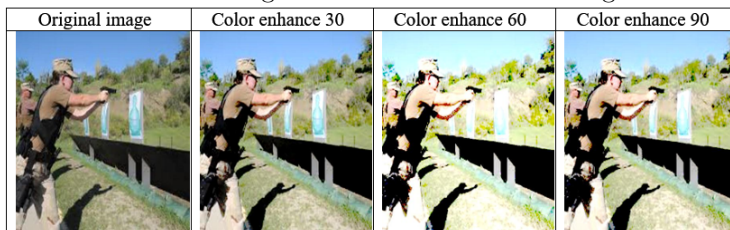
Table 2: Original and color enhanced images

Original image	Color enhance 30	Color enhance 60	Color enhance 90

features the neural net was exposed to. PCA compresses the data and retains principle features in the original image[20][8]. The PCA works substantially better for binary images that has only have two dimensions as opposed to three dimensions of RGB images [13]. While there are ways to perform PCA for RGB images using multi-linear subspace learning algorithms[13], performing PCA on binary images often produces cleaner results. However, binary PCA reduced images will do not take on color as the principle components are analyzed as 1 and 0 inputs. To restore the original color to the PCA image, MATLAB was used to fuse the original image with the PCA binary reduction image. The resultant color scheme differs from the original image because of the merging process. However, most of the color features are retained through this method.

Table 3 shows the comparison between the original image, PCA reduced image and fused image used for analysis. Figure 4 shows the performance of the VGG16 net trained with PCA reduced images on unmodified test images. Figure 4 also shows the performance of the PCA trained net on PCA reduced test images. From Figure 4, it can be observed that the PCA reduced detector tested with the original testing dataset resulted in a significant decrease in average precision from 0.2138 to 0.1223. In comparison, the PCA reduced detector tested with PCA reduced testing dataset resulted in a less significant

Figure 3: Color enhancement detector performance

decrease in average precision from 0.2138 to 0.1636. This disparity between the test results can be attributed to the features learned from the PCA reduced images failed to transfer to normal images.
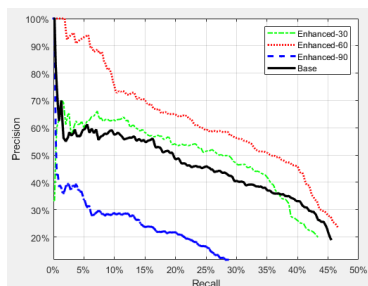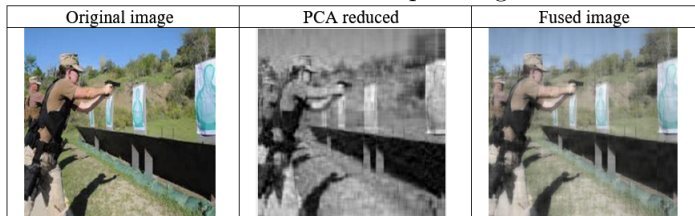
Table 3: PCA sample images

Original image	PCA reduced	Fused image

7   Combined Dataset

In order for a neural network to retain information about an object from various angles and lighting conditions, a great variance in the training dataset is required[4]. While individual approaches can yield no conclusive results towards increasing accuracy, combining the various approaches into a single dataset can increase data variations without requiring additional images. The first combined dataset used a combination of the color enhance-30, color enhance-60, PCA analysis and original images to form the 1500 training dataset. The second combined dataset used a combination of the three color enhance trials and original images. Since color enhance-60 yielded the best results, the combined dataset

Figure 4: PCA detector performance

uses 750 images from the color enhance-60 dataset, 250 images each from color enhance-90, color enhance-30, and the original images.Figure 5 shows the combined F-RCNN performance.

From Figure 5 it can be observed that combining PCA reduced images with color enhanced images had no statistically significant changes in performance as the average precision raised from 0.2138 to 0.2398. From Figure 5 it can also be observed that mixing different color enhancement ratios significantly decreases the average precision to 0.1524.
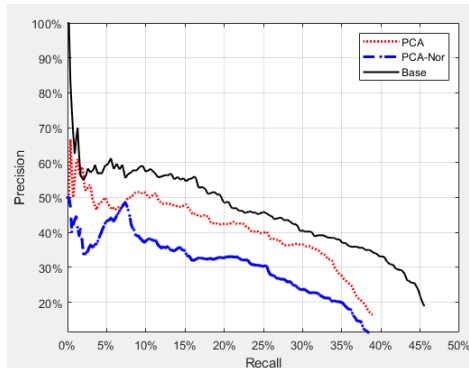
8   Discussion

Of the commonly used image preprocessing techniques in machine vision research, three methods were applied to address shortcomings of the base detector. Color enhancement was chosen to address the problem of target and background separation. Its flexible nature with many parameters makes it easily transferable from detecting pistols to other objects of interest. Secondly, PCA was introduced to reduce the number of overlapping features in each image thus lowering the chance for overfitting. Finally, the combined approach focused on merging the two previously used techniques to produce high variations in the features to address overfitting. Three different color enhancement ratios were used at intervals of 30, 60, and 90 with varying results. Figure 3 shows color enhancement-30 only increased the average precision by 0.007 from 0.2138 to 0.2208, which is not significant to indicate positive influence. The second trial increased the enhancement ratio to 60 and resulted in a noticeable rise in average precision by 0.0806 from 0.2138 to 0.2944. Although this result is not a significant change in precision, it proves that basic color enhancement can mitigate some of the background noises associated with low lighting environments. The third trial increased the enhancement ratio further to 90 and resulted in a significant drop in average precision of -0.1379 from 0.2138 to 0.0759. From the color enhancement trials it can be determined that the ratios of enhancement greatly effect the detector performance; too little color enhancement results in no significant changes but a huge enhancement ratio can have negative impacts. The rise in precision with the color enhancement-60 trial can be attributed to correctly separating key features of the pistol from the background while retaining most of the context. The color enhancement-90 trial likely overly enhanced the image so that key features of the pistol might be lost while the background context became too mono-toned.

Although the current color enhancement trials do not show significant increases in accuracy, the preliminary results obtained using minimalist preprocessing algorithms opens the way for more advanced enhancement techniques. For future works, potential extensions of this algorithm are directly amplifying a range of pistol color spectrums with machine learning algorithms.

Applying PCA to the training dataset did not result in any increases in average precision. PCA reduced the number of features by half in the training data images, which resulted in a loss of distinguishing features for the VGG16 to use. Figure 4 showed that the PCA trained detector performed poorly when tested on the original testing dataset as the average precision dropped from 0.2138 to 0.1223. In comparison, the detector performed noticeably better when tested on the PCA reduced testing dataset with the average precision only dropping from 0.2138 to 0.1636. This change in performance can be attributed to the VGG16 only retaining features pertinent to the PCA images. While too many features can confuse the neural network, too little feature will cause the learned PCA features not transferring to original images.

Figure 5: Combined detector performance

The first combined dataset with a evenly distributed color enhanced, PCA and original images only raised the average precision from 0.2138 to 0.2398. The insignificant change in average precision indicates that blind variations in training dataset does not necessarily improve performance due to conflicting features. The PCA reduced features likely interfered with the positive effects of the color enhanced trials. The second dataset focused entirely on the effect of varying color enhancements. The second combined dataset hoped to retain the positive effects from the color enhance-60 trials while adding beneficial contrast variations. Contradicting initial predictions, the second detector trained with the three color enhanced image dataset performed drastically worse than the original detector. The combined color enhanced detector lowered the average precision from 0.2138 to 0.1524. The significant drop in average precision can be attributed to the color enhance-30 and 90 dataset offering conflicting features that overruled the positive effects of the color enhance-60 dataset.

9   Conclusion

In conclusion, this paper analyzed three different image preprocessing techniques that can be applied to increase the performance of a VGG16 F-RCNN network. The color enhancement method proved to be most impactful in increasing the average precision of F-RCNN detector. Small amount of color enhancement is unlikely to result in noticeable changes in detector performance but too much enhancement have negative impacts on performance. Finding the right color enhancement ratio will depend on the target object and the relative background. It is recommended that different object datasets go through extensive trials to find the correct enhancement ratios to increase precision. Applying PCA to reduce the amount of features in the image proved to be an insufficient method by its self. The F-RCNN detector trained with PCA reduced images did not transfer the learned features to original images. While PCA reduced images by itself did not have any positive effect on the average testing precision, by combining
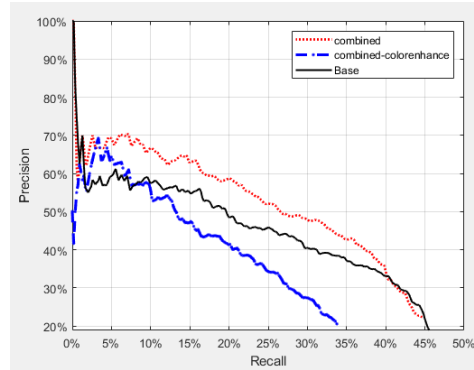
PCA reduced images with color enhanced images the detector performed on par with the base detector. However, too much dataset variations can also result in conflicting features and have a net negative impact on performance as indicated by the combined color enhancement trial.

10   Acknowledgements

References

[1] Arash Akbarinia and Karl R Gegenfurtner. How is contrast encoded in deep neural networks? *arXiv preprint arXiv:1809.01438*, 2018.

[2] Samet Akcay, Mikolaj E Kundegorski, Chris G Willcocks, and Toby P Breckon. Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery. *IEEE transactions on information forensics and security*, 13(9):2203–2215, 2018.

[3] Md Zahangir Alom, Tarek M Taha, Christopher Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Brian C Van Esesn, Abdul A S Awwal, and Vijayan K Asari. The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv preprint arXiv:1803.01164*, 2018.

[4] Christopher M Bishop et al. *Neural networks for pattern recognition*. Oxford university press, 1995.

[5] Alberto Castillo, Siham Tabik, Francisco Pérez, Roberto Olmos, and Francisco Herrera. Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning. *Neurocomputing*, 330:151–161, 2019.

[6] Atul Dhingra. Model complexity-accuracy trade-off for a convolutional neural network. *arXiv preprint arXiv:1705.03338*, 2017.

[7] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[8] Ian T Jolliffe. Springer series in statistics. *Principal component analysis*, 29, 2002.

[9] Phil Kim. Matlab deep learning. *With Machine Learning, Neural Networks and Artificial Intelligence*, 130, 2017.

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[11] Jianan Li, Xiaodan Liang, ShengMei Shen, Tingfa Xu, Jiashi Feng, and Shuicheng Yan. Scale-aware fast r-cnn for pedestrian detection. *IEEE transactions on Multimedia*, 20(4):985–996, 2017.

[12] M Lokanath, K Sai Kumar, and E Sanath Keerthi. Accurate object classification and detection by faster-rcnn. In *Materials Science and Engineering Conference Series*, volume 263, page 052028, 2017.

[13] Haiping Lu, Konstantinos N Plataniotis, and Anastasios N Venetsanopoulos. A survey of multilinear subspace learning for tensor data. *Pattern Recognition*, 44(7):1540–1551, 2011.

[14] Alasdair McAndrew. An introduction to digital image processing with matlab notes for scm2511 image processing. *School of Computer Science and Mathematics, Victoria University of Technology*, 264(1):1–264, 2004.

[15] N.A. Imfdb: Internet movie firearms database., 2020. [Accessed on 5 May 2020].

[16] James G Nagy, Katrina Palmer, and Lisa Perrone. Iterative methods for image deblurring: a matlab object-oriented approach. *Numerical Algorithms*, 36(1):73–93, 2004.

[17] Roberto Olmos, Siham Tabik, and Francisco Herrera. Automatic handgun detection alarm in videos using deep learning. *Neurocomputing*, 275:66–72, 2018.

[18] Jean Ponce, Tamara L Berg, Mark Everingham, David A Forsyth, Martial Hebert, Svetlana Lazebnik, Marcin Marszalek, Cordelia Schmid, Bryan C Russell, Antonio Torralba, et al. Dataset issues in object recognition. In *Toward category-level object recognition*, pages 29–48. Springer, 2006.

[19] Amjad Rehman and Tanzila Saba. Neural networks for document image preprocessing: state of the art. *Artificial Intelligence Review*, 42(2):253–273, 2014.

[20] Gregory Shakhnarovich and Baback Moghaddam. Face recognition in subspaces. In *Handbook of face recognition*, pages 141–168. Springer, 2005.

[21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[22] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.

[23] David PBTB Strik, Alexander M Domnanovich, Loredana Zani, Rudolf Braun, and Peter Holubar. Prediction of trace compounds in biogas from anaerobic digestion using the matlab neural network toolbox. *Environmental Modelling & Software*, 20(6):803–810, 2005.

[24] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[25] Gyanendra K Verma and Anamika Dhillon. A handheld gun detection using faster r-cnn deep learning. In *Proceedings of the 7th International Conference on Computer and Communication Technology*, pages 84–88, 2017.

[26] Yong Xu, Zheng Zhang, Guangming Lu, and Jian Yang. Approximately symmetrical face images for image preprocessing in face recognition and sparse representation based classification. *Pattern Recognition*, 54:68–82, 2016.

[27] L Yann, H Fu, and B Leon. Learning methods for generic object recognition with invariance to pose and lighting: Computer vision and pattern recognition. cvpr 2004. In *Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, 2004.

[28] Yiding Zhu, Huijing Yuan, Chuanhong Zhang, and Cunbiao Lee. Image-preprocessing method for near-wall particle image velocimetry (piv) image interrogation with very large in-plane displacement. *Measurement Science and Technology*, 24(12):125302, 2013.