University of Alberta

Computing Science

Literature Review
Developments in Artificial Intelligence

Related Works

In 2011, Bastan et al. [2] proposed the Bag-of-Visual-Words on Baggage X-Ray Images and concluded that straight forward application of BoW on X-ray Imagesdoesnot perform well as it does on regular images. 2013,Turcsany et al [10] proposed classification technique for object detection in X-ray baggage imagery using primed visual words in an SVM classifier framework. Primed visual words are obtained through class-specific clustering of fea-ture descriptors and used to encode images in bag-of-words model. This differs from the traditional approach, which combines the feature set of positive and negative classes during the clustering process when generating a codebook. Modification to the clustering stage of the traditional bag-of-words framework creates an image representation scheme that further facilitates the separation of positive and negative class.This method significantly outperforms the previous work of Bastan et al. [2]. People start using CNN networks for Object Detection in X-ray images.Akcay et al [1] ex-amines the applicability of traditional sliding window convolutional neural network (CNN) detection and the relative performance of contemporary object detection strategies forregion based object detection techniques – Faster Region-based CNN (R-CNN) and Region Based Fully ConvolutionNetworks(R-FCN) on X-ray securities images. Use Transfer learning due to limitation of training dataset of 11,627samples (5,867 training, 2,880 validation and 2,880 test samples). The Faster RCNN and R-FCNN provide superior results than traditional sliding window driven CNN(SW-CNN) approach.Faster RCNN with VGG16, pretrained onthe ImageNet dataset, achieved 88.3 mAP for a six object -firearm, firearm-components, knives, ceramic knives, camera and laptop detection in X-ray dataset. R-FCN with ResNet-101, yields 96.3 mAP for the two class firearm detection problem and requires 100 milli second computation per image

Model	Network	mAP	camera	laptop	gun	gun component	knife	ceramic knife
SWCNN	AlexNet	60.8	68.2	60.9	74.8	71.4	21.2	68.3
	VGGM	63.4	70.7	63.7	76.3	73.1	24.6	71.9
	VGG16	64.9	70.1	72.4	75.2	75.7	22.3	73.4
	ResNet-50	67.1	69.2	80.1	74.7	76.1	31.4	71.3
	ResNet-101	77.6	88.1	90.2	83.1	84.8	39.2	80.3
RCNN	AlexNet	64.7	79.1	81.5	85.3	58.2	18.8	65.8
	VGGM	68.6	79.9	85.5	86.9	65.8	21.0	72.3
	VGG16	77.9	88.8	95.4	87.6	83.2	30.4	81.9
FRCNN	AlexNet	78.8	89.3	75.6	91.4	87.4	46.7	82.3
	VGGM	82.3	90.0	83.4	91.8	87.5	54.2	86.9
	VGG16	**88.3**	88.1	91.8	92.7	93.8	72.1	91.2
	ResNet-50	85.1	84.4	87.9	91.6	90.1	67.7	88.9
	ResNet-101	87.4	85.7	90.4	93.1	91.1	73.2	90.7
RFCN	ResNet-50	84.6	89.4	92.8	93.2	91.8	50.6	89.6
	ResNet-101	85.6	88.7	90.6	94.2	92.5	55.6	92.0

Figure 1: Detection results of SW-CNN, Fast-RCNN (RCNN) , Faster RCNN (FRCNN) and R-FCN for multi-class problem (300 region proposals) [3]

Petrozziello et.al [4] compare several algorithms for detection of firearm parts(steel barrel bores) in x-ray images of travelers' baggage. Dataset of 22k double view x-ray scans, containing a mixture of benign and threat objects was used. During preprocessing, a smoothing algorithm is applied to the grayscale images in order to reduce the low-level noise within it. Then, a thresholding algorithm followed by normalization was applied such that metal objects, such as firearm barrels, and other components are kept as shown in the image below.

Figure 2 A sample image containing a steel barrel bores (top right cylinder in the top row) from the parcel dataset. The left image (both rows) is the raw dual view x-ray scan, in the middle, the grey scale smoothed one, and on the right, b/w thresholded one. The parcel dataset usually contains a higher amount of steel objects and the barrels are better concealed.

Out of Convolutional Neural Networks (CNN) , Stacked Autoencoders, Neural Networks, and Random Forests CNNs outperformed all the other methods.

Bhowmik et al. [1] discusses on how to synthetically cre-ate x-ray images with prohibited items(Data augmentation).This creates an opportunity for as the existing public domaindatasets such as GDXray [3] contains lesser clutter ,overlapand contains limited categories. SIXray [4] has dataset withclutter but has smaller number of categories

Table 1 Baggage dataset results for the AUC, FPR@90%TPR and F1-Score metrics. The results are reported for the four classification techniques and three pre-processing step: raw data, grey scale smoothing and b/w thresholding.

Metric	Technique	Raw	Smoothing	B/w thresholding
AUC	CNN	**93**	**95**	**96**
	Autoencoder	75	78	90
	oBIFs + NN	85	87	94
	oBIFs + RF	66	72	80
FPR @ 90% TPR	CNN	**9**	**7**	**6**
	Autoencoder	70	60	26
	oBIFs + NN	50	31	14
	oBIFs + RF	86	66	53
F1-Score	CNN	**91**	**93**	**93**
	Autoencoder	60	65	81
	oBIFs + NN	64	67	79
	oBIFs + RF	36	41	56

Samet Akçay[5], proposed transfer Learning using convolution neural networks for object classification within X-Ray Baggage Security Imaginary. Discusses various CNN-based methods (GoogLeNet & AlexNet) with Transfer Learning for feature extraction, representation, and classification process. This seems to be the first work on automatic X-ray screening with DCNN. This displays that CNN is able to achieve better performance when compared to other methods.

2020, Hassan et.al [5] proposed a novel method to overcome the retraining requirement of framework across multiple scanner-specification. It uses meta-transfer learning-driven tensor shot detector that decomposes the candidate scan into dual-energy tensors and employs a meta-one-shot classification backbone to recognize and localize the cluttered baggage threats. This method can be well generalized for multiple scanner specifications due to its capacity to generate object proposals from the unified tensor maps rather than diversified raw scans. On the SIXray dataset,the proposed framework achieved a mean average precision(mAP) of 0.6457, and on the GDXray dataset, it achieved the precision and F1 score of 0.9441 and 0.9598, respectively.

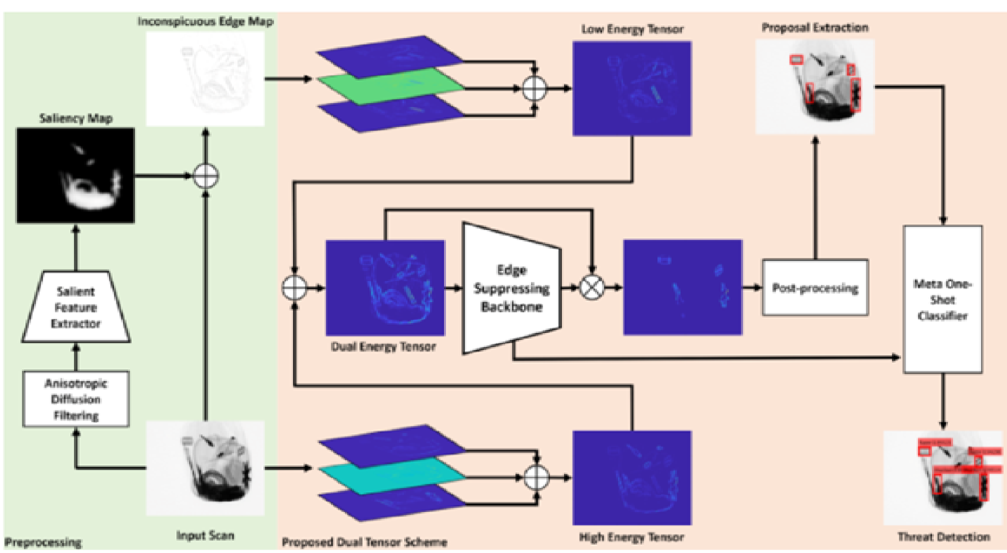

Figure 2. Block diagram of the proposed tensor-shot detector

List of Other Papers Reviewed [4+ 4 = 8 Papers]

Source: “Multimedia Book ”

1. Semantic Learning For Image Compression

This work presents a novel architecture developed specifically for image compression, which generates a semantic map for salient regions so that they can be encoded at a higher quality as compared to background regions. For the MS-ROI model, the architecture of the CAM model is modified in order to obtain heat maps that allow for the localization of multiple regions of interest. Used modified ResNet-50 architecture for identifying the region of interest. The efficiency of JPEG compression is based on the spatial frequency, or concentration of detail, of the image. Areas in the image with low frequency are compressed more than areas with higher frequency. This characteristic of JPEG along with this approach helps to outperform standard JPEG in terms of compression while preserving the latent semantics in the compressed image.

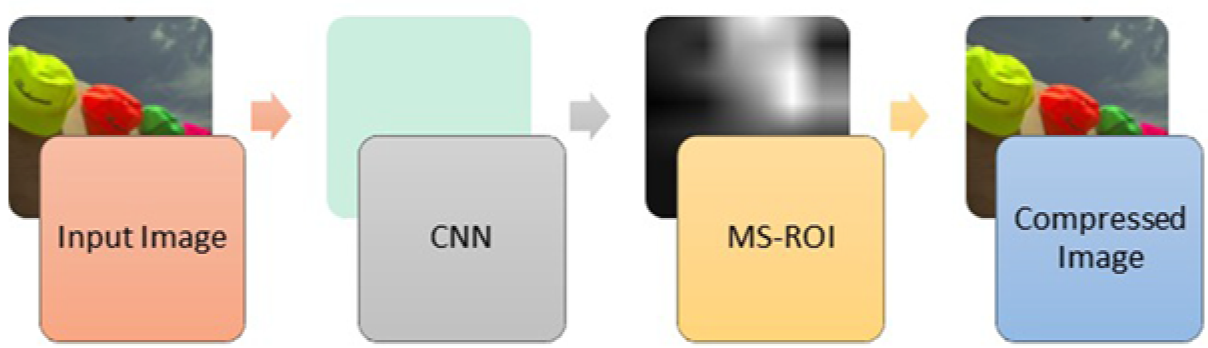

Fig. 1. Basic flow diagram of the developed approach

2. ProDeblurGAN: Progressive Growing of GANs for Blind Motion Deblurring in Face Recognition

NVidia introduced progressively growing GAN (ProGAN) in 2017 to address some of the issues in stabilizing the GAN training and producing high-resolution images up to 1024×1024 resolution. This paper presents ProDeblurGAN an expansion of the ProGAN, improvised and tailor-made to perform deblurring of a single image. Gaussian kernel-based blur was applied to all the sharp images to prepare the dataset. Different models are compared on similar datasets and state of art metrics.

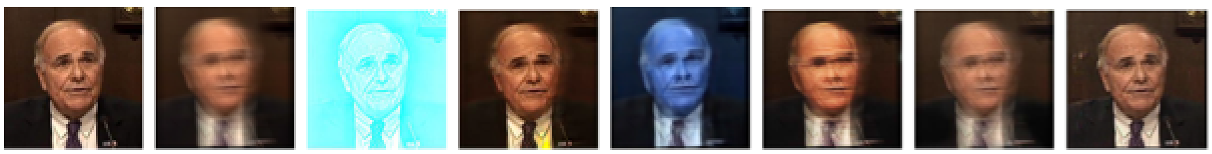

Fig3: From left to right, original sharp image, blurred image input, deblurred image using SRN (result after 10 epoch), deblurred image using SRN (best result after 5 epoch), deblurred image using DeblurGAN GoPRO, deblurred image using DeblurGAN VGGFace2, deblurred image using ProDeblurGAN

3. Homography-Based Vehicle Pose Estimation from a Single Image by Using Machine-Learning for Wheel-Region and Tire-Road Contact Point Detection

The paper discusses an image-based photogrammetry system for measuring vehicle lane poses using a single perspective camera where a Deep-Learning-based technique(YOLOV3) is developed for identifying/classifying the wheels on a vehicle, as Regions of Interests, and extracting the tire-road contact point from the image, and using a Homography-based approach to extract vehicle pose.

For finding the region of interest, the algorithm takes perspective images as input, applies homography, and produces a side view of the scene by mapping the pixels to an orthogonal coordinate
frame

a b

Fig. 3. (a) original image, (b) detected tire-road contact point in the perspective image.

Fig. 6. Results of position estimates using tire-road contact point.

The image dataset consists of 2104 images labeled by experts which were captured by different inspection cameras, and under different lighting and weather conditions.

4. Human Body Fall Recognition System

The paper discusses a fall or n-fall classification which detects and creates an alert in real-time by using OpenCV with human detection algorithms. The paper uses a Le2I data-set which includes various scenarios of human activities and the environment
GMG and MOG2 algorithms are used for background separation in each .frames. Then blur and averaging filters are applied to each frame, here we have a binary mask frame. Now the human body is approximated with a bounding box and an ellipse.

The following two methods are used for fall detection.

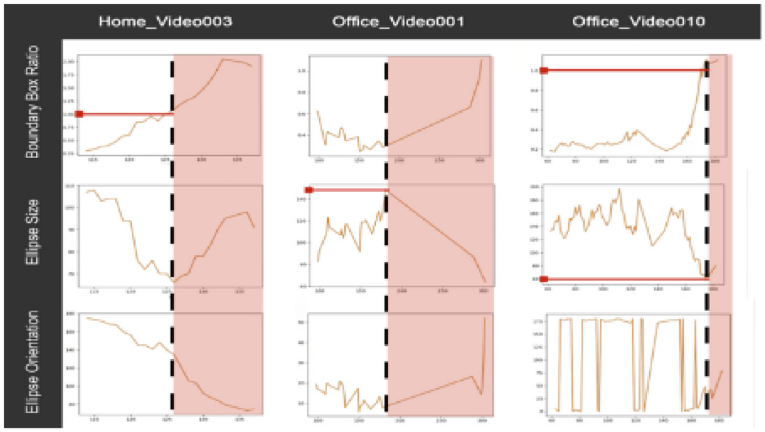

Fig. 3. Fall detection by value analysis

- Change of w/h: In this case as the size of the human body approximation changes from vertical to horizontal (w/h greater than 1), the fall is detected. (subject falls to sides)
- Change of θ and f: As θ gets closer to π or 0 degrees, the change in the angle is used to detect the fall. (the subject falls towards the camera)

Source: “MRC Papers”

1. Gaussian and Laplacian of Gaussian weighting functions for robust feature based tracking:

This paper addresses the effects of noise on the performance of the standard KLT algorithm and proposes two solutions - Laplacian of Gaussian (LoG) and Gaussian weighting function to increase immunity of the tracking algorithm to effects such as poor CCD exposure and compression noise. The KLT algorithm is limited by its use of averaging as a weighting function over the feature window.. KLT algorithm fails to track the object even at low levels of noise, while the weighted tracking algorithm continues to track until 25% noise. The LoG is suitable for images with well-defined and sharp image corners like the synthetic test images. Real images where the edges may not be as sharply defined, the Gaussian weighting function performs better.

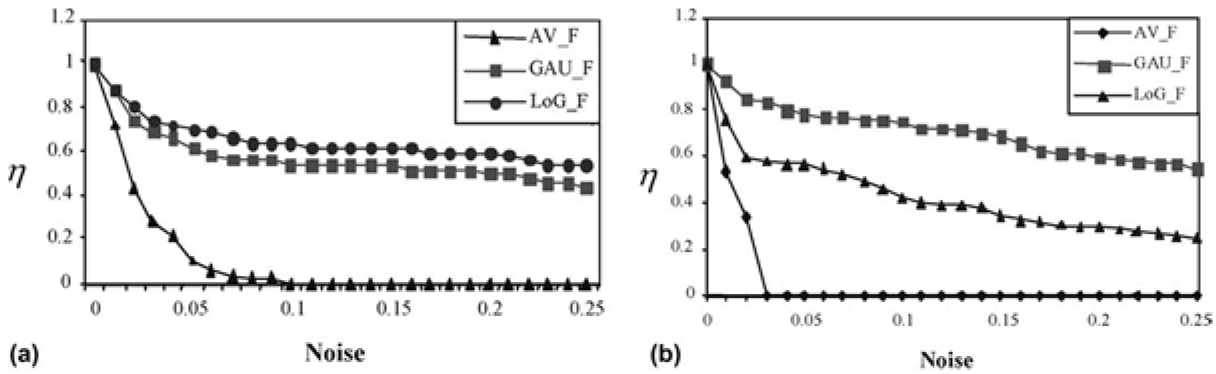

Fig. (a) Performance of weighting functions with synthetic car sequence, 50 features have been initialized and are being tracked. (b) Performance of weighting functions with real image car sequence, 50 features have been initialized and are being tracked.

2. A Framework for Adaptive Training and Games in Virtual Reality Rehabilitation Environments:

Bayesian networks as a way to measure student performance in online web-based multimedia educational games but have not been used in Virtual Reality Rehabilitation.This paper proposed a novel framework based on Bayesian networks for self-adjusting adaptive training in virtual rehabilitation environments. Participants who used the proposed virtual reality wheelchair training system completed the real world obstacle course on average faster, with a mean time of 81.5 seconds for the experimental group compared to 104.5 seconds for the control group .

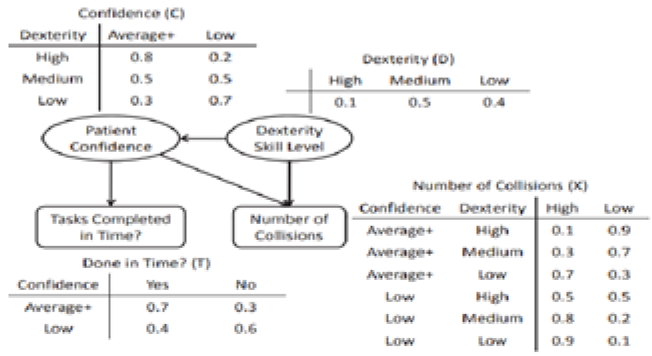

Figure: A sample Bayesian network topology

3. Integrating active face tracking with model-based coding

This paper discusses the tracking of a talking face with an active camera whereas most previous work done focuses on the detection of a still camera. The detection and tracking of the active talking face is a fundamental step towards realizing an application of MPEG4 in real situations. First, the background compensation in successive frames is done, and (Murray and Basu, 1994), the motion-energy tracking approach is used coupled with a morphological filter to reduce the noise. Third, the facial features are detected using Hough Transform and deformable template coupled with color information. Finally, a wireframe model is adapted to the extracted face using a coarse-to-fine adaptation algorithm.

4. Visual gesture recognition for ground air traffic control using the Radon transform

The paper discusses a novel method for the recognition of hand gestures, used by air marshals for steering aircraft on the runway, using the Radon transform. Foreground is extracted from background by using an adaptive threshold value. Thinning is applied to images to obtain the medial skeleton

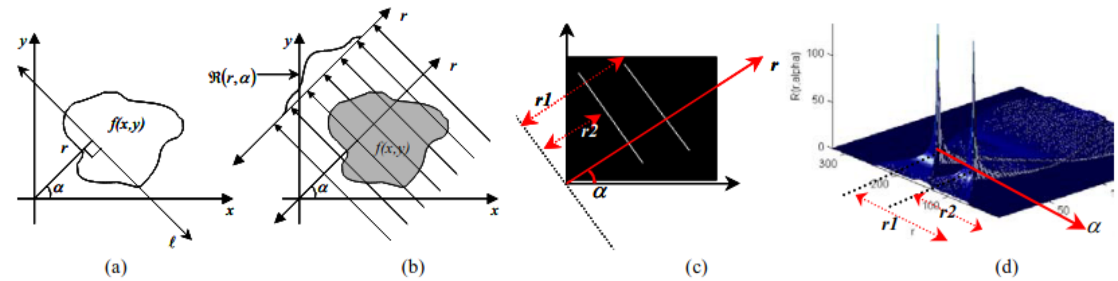

Fig.1: (a) Pictorial representation of line used in (1), (b) Illustration of RT as an integration of the image; RT of two parallel lines - (c) Skeleton image, (d) Mesh plot of RT of parallel lines in (c).

Now Radon Transform is applied to detect the lines as shown below :

Fig.4: Sample images showing (a) original data, (b) binary image data and (c) results of the medial skeletonization operation.

Once the Radon transform has been computed, we threshold the Radon transform coefficients to extract the most significant local regions. It is then normalized between 0-1. Finally, a k means clustering with k=8 is applied for classification.

References

[1] Samet Akc ̧ay, Mikolaj E Kundegorski, Michael Devereux, and Toby PBreckon. Transfer learning using convolutional neural networks forobject classification within x-ray baggage security imagery. In2016IEEE International Conference on Image Processing (ICIP), pages1057–1061. IEEE, 2016.

[2] Muhammet Bas ̧tan, Mohammad Reza Yousefi, and Thomas M Breuel.Visual words on baggage x-ray images. In International Confer-ence on Computer Analysis of Images and Patterns, pages 360–368.Springer, 2011

[3] Neelanjan Bhowmik, Qian Wang, Yona Falinie A Gaus, Marcin Szarek, and Toby P Breckon. The good, the bad and the ugly: Evaluating Convolutional neural networks for prohibited item detection using real and synthetically composited x-ray imagery.arXiv preprintarXiv:1909.11508, 2019.

[4] Victoria Cutler and Susan Paddock. Use of threat image projection (tip)to enhance security performance. In the 43rd Annual 2009 InternationalCarnahan Conference on Security Technology, pages 46–51. IEEE,2009

[5] Taimur Hassan, Muhammad Shafay, Samet Akc ̧ay, Salman Khan,Mohammed Bennamoun, Ernesto Damiani, and Naoufel Werghi.Meta-transfer learning driven tensor-shot detector for the autonomouslocalization and recognition of concealed baggage threats.Sensors,20(22):6450, 2020.

[6] Domingo Mery, Vladimir Riffo, Uwe Zscherpel, German Mondrag ́on,Iv ́an Lillo, Irene Zuccar, Hans Lobel, and Miguel Carrasco. Gdxray: The database of x-ray images for nondestructive testing.Journal of Nondestructive Evaluation, 34(4):1–12, 2015

[7] Caijing Miao, Lingxi Xie, Fang Wan, Chi Su, Hongye Liu, JianbinJiao, and Qixiang Ye. Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images. InProceedings of the IEEE/CVF Conference on Computer Vision and pattern Recognition, pages 2119–2128, 2019

[8] Stefan Michel, Saskia M Koller, Jaap C de Ruiter, Robert Moerland, Maarten Hogervorst, and Adrian Schwaninger. Computer-based training increases efficiency in x-ray image interpretation by aviation security screeners. In2007 41st Annual IEEE international Carn

[9] Alessio Petrozziello and Ivan Jordanov.Automated deep learning for threat detection in luggage from x-ray images. InInternationalSymposium on Experimental Algorithms pages 505–512. Springer,2019.

[10] Diana Turcsany, Andre Mouton, and Toby P Breckon. Improvingfeature-based object recognition for x-ray baggage security screeningusing primed visualwords. In2013 IEEE International conference onindustrial technology (ICIT), pages 1140–1145. IEEE, 2013.

[11] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object detection with deep learning: A review.IEEE transactions on neural networks and learning systems, 30(11):3212–3232, 2019.