Scientific Research Publishing

Advances in Molecular Imaging

www.scirp.org/journa/ami

Table of Contents

Volume 8   Number 4   October 2018

Advances in Molecular Imaging (AMI)
Journal Information

COPYRIGHT

Human Face Super-Resolution Based on Hybrid Algorithm

Jinfeng Xia, Zhizheng Yang, Fang Li, Yuanda Xu, Nan Ma, Chunxing Wang*

Shandong Normal University, Jinan, China
Email: *cxwang@sdnu.edu.cn

Abstract

Aiming at the problems of image super-resolution algorithm with many convolutional neural networks, such as large parameters, large computational complexity and blurred image texture, we propose a new algorithm model. The classical convolutional neural network is improved, the convolution kernel size is adjusted, and the parameters are reduced; the pooling layer is added to reduce the dimension. Reduced computational complexity, increased learning rate, and reduced training time. The iterative back-projection algorithm is combined with the convolutional neural network to create a new algorithm model. The experimental results show that compared with the traditional facial illusion method, the proposed method can obtain better performance.

Keywords

Face Hallucination Super Resolution, Convolutional Network Hybrid Algorithm

1. Introduction

Image super-resolution is a classical problem in the domain of computer vision. It aims to infer an HR image with crucial information from the given LR images. Face hallucination is a branch of image super-resolution, which develops domain specific prior knowledge with strong cohesion to face domain. It was first introduced by Baker and Kanada [1] and has attracted growing attention due to practical importance in many face based applications such as face recognition, face alignment and so on. As the development of machine learning, there are numerous learning-based methods which have been proposed to solve the face hallucination problem. Learning based algorithms have been seen to achieve higher magnification factor with better visual quality than the other super reso-

lution techniques such as bi-cubic interpolation and reconstruction based techniques.

The algorithm based on Interpolation is the most basic method in face super-resolution research, including the nearest neighbor interpolation, bilinear interpolation, bicubic Interpolation etc. The method based on reconstruction has a fast speed and made a little improvement in image quality. However, because it is limited by the original information of the image, the ambiguity caused by low resolution sampling cannot be removed. Feerman *et al.* [2], proposed a method of image super-resolution based on sample learning, and obtained the vector set in the external database through the nearest neighbor search. Timofte combined the sparse learning dictionary and the domain embedding method, and proposed the method of fixed domain regression, which improves the processing speed of the algorithm. Dong *et al.* [3], proposed a deep convolutional neural network (CNN) based single image super resolution method and showed that the traditional sparse coding-based algorithm can also be seen as a kind of deep convolutional network. The end-to-end mapping between LR images and HR images was optimized in Dong's SR method, which achieves excellent reconstruction performance.

Inspired by the above literature, we apply the deep learning theory to illusory face hallucination reconstruction [4]. The model of deep convolution neural network is improved, the convolution neural network is added to the pool layer, the convolution kernel size is adjusted, the parameters are reduced, and the operation speed is increased. Finally, the iterative back projection method is used to reconstruct the face image after post-processing [5].

2. Methodology

Image acquisition process may be affected by motion blur, optical blur, signal aliasing caused by down sampling and all kinds of noise. The picture is polluted by all of the above. Elad proposed a matrix vector approach to describe low resolution image imaging models [2].

$$y = HDXw + n \qquad (1)$$

X represents high resolution images, y means low resolution images, and N represents additive Gauss noise. D, H and w denote the down sampling matrix, the fuzzy matrix and the geometric transformation matrix respectively. Hallucination face is the inverse process of face image degradation. The purpose is to give the low resolution image y to restore the original high resolution image X.

Convolutional Neural Networks are a biologically inspired variant of multi-layered perceptron networks (MLP's), specialized for image processing. First popularized by LeCun *et al.* in they are similar to other hierarchical feature extraction methods such as the Neocognitron and HMAX.

The structure of a typical CNN consists of alternating layers of convolutional and pooling followed by an output classification layer. Each type of layer contains several feature maps, or groups of neurons, in a rectangular configuration..

The receptive field itself is simply a number of weighted connections, that is, each connecting edge has a weight. The group of weights applied by a neuron is called a weight kernel. A distinguishing property of these networks is that all neurons in a feature map share the same weight kernel. The idea behind this configuration is that a spatial feature detector should be useful across an entire image, instead of just at a particular location; for example a vertical edge detector. The convolutional layers in the network perform the majority of processing in these networks, with the feature maps in the pooling layers simply down-sampling their corresponding feature map in the convolutional layer [6].

In the past few years, the method based on deep learning has been improved and updated. It is not only applied to the image classification in the field of computer vision, but also from face recognition to semantic segmentation. Recently, deep learning method has also been applied to low level vision tasks, including image denoising, image enhancement, image super-resolution and so on. The seminal work of image super-resolution convolutional neural network (SRCNN) was done by Dong *et al.* [7].

The model is mainly composed of three volumes, which generally simulate a sparse layer. Three coiling layers accomplish the following tasks: patch extraction and representation, nonlinear mapping and reconstruction [8].

1) Patch extraction and representation: A patch is extracted from a low resolution image and each patch is represented as a high dimensional vector. These high dimensional vectors are composed of a set of feature maps, and the dimension of the vector is equal to the number of maps.

2) Non-linear mapping: this step nonlinearly maps each high dimensional vector onto another high dimensional vector. Each mapped vector is conceptually the representation of a HR patch. These vectors comprise another set of feature maps.

3) Reconstruction: this step aggregates the above HR patch-wise representations to generate the final HR image. This image is expected to be similar to the ground truth R.

2.1. Iterative Back Projection Algorithm

The iterative back projection algorithm proposed by Irani is the representative method of the original image restoration [4]. It is used as the post-processing in the processing of the image super-resolution algorithm. The result of the single image interpolation is usually used as the initial solution of the high resolution image. According to the system model, the quasi low resolution image can be expressed as:

$$y_0 = Hx_0 + n \tag{2}$$

If the x is equal to the original high resolution image and the upper analog imaging process conforms to the actual situation, the analog low resolution sequence y_0 is the same as the actual low resolution image y, and if it is different, the difference between y and y_0 is projected back to x_0 for correction.

In IBP algorithm, HR image is obtained by utilizing the backward projection of the error projection matrix based on the difference between simulated LR images and the observed LR images with up-sampling, reverse blur filter and reverse motion transform [5] [9]. This process iteratively calculates the estimated HR image until the energy of the error comes to a minimum value, or the iteration step achieves the maximum number. The simulated LR images, the error projection matrix for the acquisition process, and the IBP algorithm iterative reconstruction process can be written as:

$$E^n = \frac{1}{N} N \sum_{K=1}^{N} H_k^{BP} \left(y_{k-} \hat{y}_k^{(n)} \right) \tag{3}$$

$$\hat{z}^{(n+1)} = \hat{z}^{(n)} + \lambda E^{(n)} \tag{4}$$

where $\hat{z}$ and $\hat{z}^{(n+1)}$ denote the super resolution image gained from the (n)th and ($n + 1$)th iteration respectively, $\hat{y}_k^{(n)}$ denotes the (n)th simulated LR images of $\hat{z}^{(n)}$ under the imaging degradation model, $E^{(n)}$ is the difference between the simulated LR images and the observed LR images, H_k^{BP} is the (k)th back projection operation and λ is the iteration step.

2.2. The Hybrid Algorithm

The reconstruction algorithm of the iterative back-projection algorithm is not outstanding, but it can be combined with other super resolution methods to improve the performance. In this paper, the super-resolution algorithm based on convolution neural network is improved and combined with the iterative back-projection algorithm [10], a new composite algorithm is proposed.

2.2.1. The Convolution Layer

In the convolution layer, we mainly consider the influence of the size and number of convolution kernel on the processing effect and processing speed of the model, Convolution neural network proposed by Dong *et al.* [3] [7]. And super-resolution model have three volumes of convolution kernel size of 9 × 9, 1 × 1, 5 × 5. The improved algorithm in this paper has a three layer convolution kernel size of 3 × 3, 1 × 1, 3 × 3. As shown in the following Figure 1.

The larger the convolution kernel size is in convolution, the better the super-resolution effect, but it will also increase the corresponding computation [11]. In our improved convolutional neural network model, the first layer convolution kernel is changed to 3 × 3, which can effectively reduce the number of parameters and cover the features of the image. If the second layer convolution kernel increases, the parameter will increase. The convolution kernel size of second layers is still 1 × 1. The size of the convolution kernel of third layers is 3 × 3. As the number of convolution cores may affect the quality of super-resolution images, we train and test the number of iterations of the convolution neural network model of 3-1-3 and 3-1-5 respectively. The experimental results show that each iteration 100 times 3-1-5 model takes 7.1 seconds more than the 3-1-3 model; The average PSNR 3-1-5 model of the phase output super
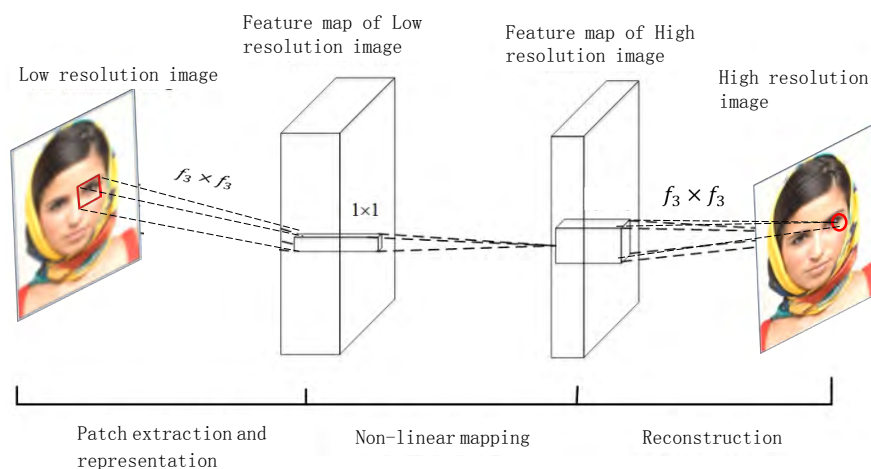
Figure 1. The flow chart of super-resolution algorithm.

resolution image is 0.1dB higher than that of the 3-1-3 model. Considering the processing results and computing speed, we select the 3-1-3 model.

The number of convolution kernel and convolution kernel size determines the super resolution effect together. In the super-resolution algorithm 9-1-5 model proposed by Dong and others, the first convolution kernel number and the second layer convolution kernel number improvement algorithm are tested in the 3-1-3 model for different size and the final selection $n_1 = 64, n_2 = 64$.

According to the selection of the number of convolution kernel [10] we chose and the test image under the condition of the same iteration number and the same learning rate in the improved algorithm 3-1-3 model. The PSNR value of super resolution image test is 5×10^4 and the learning rate is 10^{-3}. The comparison data obtained are shown in Table 1.

At the same time, the influence of the amount of calculation and information on the speed of the image super-resolution is considered. The size of the patch is selected to be 33 × 33, and the size of the sub image is properly increased. Experimental results show that the increase of input block size improves training speed and shortens training time.

2.2.2. Pooled Layer

Dong *et al.* proposed to apply convolution neural network model to image super-resolution processing, there is no pool layer in the model. In addition to improving the size and number of convolution kernels, we also introduce pooled layers [12]. After introducing the pool layer into the first, second volume layer, the introduction of the pool layer can reduce the output vector, reduce the dimension and speed up the training. The phenomenon of over fitting can be avoided after adding the pool layer. After introducing the pool layer, the network depth reaches five layers, and the network structure with deeper layers is more conducive to the learning of image data. At the same time, the number of parameters can be reduced by pooling layer [13] [14].

To make the pooling unit have translation invariance, that is, after a small

Table 1. The test PSNR values of different convolution kernel numbers.

kernel number	$n_1 = 64, n_2 = 32$	$n_1 = 64, n_2 = 64$	$n_1 = 128, n_2 = 64$
PSNR	33.57	33.65	30.14

translation, the image still produces the same characteristics as before. We can choose the continuous range in the image as the pool area and only pool the features generated by the same hidden unit, so the number of the pool layer input feature graph will not change, but the size of the feature graph will be reduced. This process is actually a down sampling process [15]. It can be expressed as a formula:

$$X_j^k = f \left[\beta_j^k \text{down} \left(x_j^{k-1} + b_j^k \right) \right] \tag{5}$$

2.3. Architecture of Hybrid Algorithm

The whole model of the hybrid algorithm is divided into nine layers, including four coiling layers, two pool layers, two subsampling layers (one for lower sampling, the other for upper sampling), and the other for differential stratification [13] [16]. The detailed description of the specific functions of each layer is as follows:

1) The first five layers are the framework of the super-resolution algorithm of the convolution neural network (SRCNN model), which are mainly implemented in the five layers: the extraction and expression of the patch, the nonlinear mapping and reconstruction.

2) Down-sampling layer. This operation down-samples the image derived from the third layer. As a result, a LR version of reconstruction image is obtained.

3) Difference layer. This operation calculates the difference between down-sampling version of the original HR image and the corresponding counterpart we acquired above. The difference is treated as reconstruct-ion simulation error, and it also can be considered as a prior guidance that has been introduced

4) Up-sampling layer. This operation up-samples the simulation error to generate the simulation error of HR version.

5) Update layer. This operation performs a convolution with the above simulation error, and then the final HR image is updated based on the synthesis of the third layer's result with the convolution version of simulation error.

3. Experimental Results

Considering the test and contrast, the image degradation model we adopted is to generate low resolution images by sampling the original high-resolution images after Gauss blur. The initial high resolution image is input into the convolution neural network model of off-line training, and the final high resolution image is generated by the processing of network learning and optimization after the

processing of the model parameters [15]. In order to ensure the objectivity and rationality of the comparison, the learning results of the network model with the same iteration number are selected for comparative test. The maximum iteration number of the composite algorithm is 5×10^5 due to device constraints.

We choose a Bicubic [17] method, the NE algorithm [18], the ScSR algorithm [19] [20] and the SRCNN algorithm [21] for comparative test. We evaluate and compare the performance of our proposed models in terms of peak signal to noise ratio (PSNR) and structure similarity (SSIM). The model is tested on 500 images of BANVA dataset. Here are the visual qualities of the face hallucination results generated by our method and other competing ones in Figure 2.

Table 2 shows the comparison results in terms of PSNR and SSIM between the proposed improved method and some competing methods. The best results are high-lighted in bold. As one can see, our proposed method improves the results both in terms of PSNR and SSIM. And they all have an up-scaling factor of in our experiments. We can make some observations from the results. Obviously, the proposed modified net-work has already achieved better results

Figure 2. Results of images with different algorithms.

Table 2. The result of PSNR (dB) and SSIM on test images using different methods.

Image	BI		NE		ScSR		SRCNN		Proposed	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	27.50	0.798	28.30	0.803	29.27	0.825	30.23	0.834	31.01	0.871
2	28.91	0.837	28.36	0.844	29.89	0.874	30.59	0.882	31.87	0.912
3	28.01	0.823	28.81	0.832	30.61	0.853	31.02	0.875	32.23	0.891
4	27.04	0.812	28.36	0.829	28.69	0.843	30.22	0.862	31.35	0.886
5	28.70	0.833	29.09	0.838	29.21	0.853	31.21	0.880	31.93	0.879
6	32.48	0.897	32.56	0.898	33.45	0.905	34.00	0.914	34.08	0.917
Average	28.77	0.833	29.25	0.840	30.197	0.859	31.21	0.875	32.08	0.890

than the competing methods, which validates that introducing image prior for face hallucination works well.

4. Conclusion

In this paper, by analyzing the training process of convolutional neural networks, we have made a series of improvements to the image super-resolution algorithm based on convolutional neural networks. Compared with the traditional algorithm, the results show that the improved algorithm has better reconstruction effect, higher edge sharpness and clearer picture. Improved convolutional neural network algorithm can achieve better results with less iteration and significantly reduce training time.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Baker, S. and Kanade, T. (2000) Hallucinating Faces. *IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, 28-30 March 2000, 83-88. https://doi.org/10.1109/AFGR.2000.840616

[2] Freeman, W.T. and Pasztor, E.C. (1999) Learning Low-Level Vision. *International Conference on Computer Vision*, Kerkyra, 20-27 September 1999, 1182-1189. https://doi.org/10.1109/ICCV.1999.790414

[3] Dong, C., *et al.* (2014) Learning a Deep Convolutional Network for Image Super-Resolution. *European Conference on Computer Vision*, 184-199.

[4] Irani, M. and Peleg, S. (1990) Super Resolution from Image Sequences. 10*th International Conference on Pattern Recognition*, Atlantic City, 16-21 June 1990. https://doi.org/10.1109/ICPR.1990.119340

[5] Plank, J., Beck, M. and Elwasif, W. (1999) *IBP: The Internet Backplane Protocol,* **93**, 679-688.

[6] Lei, J., *et al.* (2018) An Image Guided Algorithm for Range Map Super-Resolution. *Society of Photo-Optical Instrumentation Engineers*, 43. https://doi.org/10.1117/12.2288074

[7] Dong, C., *et al.* (2016) Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 295-307. https://doi.org/10.1109/TPAMI.2015.2439281

[8] Zhou, S.K., Chellappa, R. and Moghaddam, B. (2004) Intra-Personal Kernel Space for Face Recognition. *IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul, 19 May 2004, 235-240. https://doi.org/10.1109/AFGR.2004.1301537

[9] Yan, W.U. (2009) One Improved Super-Reconstruction Algorithm Based on IBP Theory. *Infrared*, **12**, 11-15.

[10] Gao, H., Zeng, J. and Zhao, Y. (2016) Super-Resolution Reconstruction Algorithm Based on Adaptive Convolution Kernel Size Selection. *Applications of Digital Image Processing XXXIX*, Article ID: 997120.

[11] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *International Conference on Neural Information Processing Systems*, 1097-1105.

[12] Wang, N., *et al.* (2014) A Comprehensive Survey to Face Hallucination. *International Journal of Computer Vision*, **106**, 9-30. https://doi.org/10.1007/s11263-013-0645-9

[13] Grm, K., *et al.* (2018) Face Hallucination Using Cascaded Super-Resolution and Identity Priors.

[14] Jiang, J., *et al.* (2018) Deep CNN Denoiser and Multi-Layer Neighbor Component Embedding for Face Hallucination. *IJCAI.* https://doi.org/10.24963/ijcai.2018/107

[15] Shi, J., *et al.* (2018) Hallucinating Face Image by Regularization Models in High-Resolution Feature Space. *IEEE Transactions on Image Processing*, **27**, 2980-2995. https://doi.org/10.1109/TIP.2018.2813163

[16] Ma, X., Zhang, J. and Qi, C. (2010) Hallucinating Face by Position-Patch. *Pattern Recognition*, **43**, 2224-2236. https://doi.org/10.1016/j.patcog.2009.12.019

[17] Ruangsang, W. and Aramvith, S. (2017) Efficient Super-Resolution Algorithm Using Overlapping Bicubic Interpolation. *6th Global Conference on Consumer Electronics*, Nagoya, 24-27 October 2017, 1-2.

[18] Cao, Q.I., *et al.* (2016) Super-Resolution Algorithm of Infrared Video Image Based on Sparse Representation.

[19] Yang, J., *et al.* (2010) Image Super-Resolution via Sparse Representation. *IEEE Transactions on Image Processing*, **19**, 2861-2873. https://doi.org/10.1109/TIP.2010.2050625

[20] Wang, Y., *et al.* (2018) Improved Algorithm of Image Super Resolution Based on Residual Neural Network. *Journal of Computer Applications*.

[21] Wang, Y., *et al.* (2016) End-to-End Image Super-Resolution via Deep and Shallow Convolutional Networks.

Minimally Invaded Sentinel Lymph Node Model for the Development of Intraoperative Infrared Fluorescence Imaging

Badiane Serigne Moussa[1]*, Raes Florian[2]*, Julien Sobilo[2], Natkunarajah Sharuja[2], Lerondel Stéphanie[2], Le Pape Alain[2]

[1]UFR des Sciences de la Santé, Université Gaston BERGER, Saint-Louis, Sénégal
[2]Centre for Small Animal Imaging, PHENOMIN-TAAM CIPA, CNRS UPS44, Orléans, France
Email: semobadiane@yahoo.fr, florian.raes@cnrs-orleans.fr, julien.sobilo@cnrs-orleans.fr,
sharuja.natkunarajah@cnrs-orleans.fr, stephanie.lerondel@cnrs-orleans.fr, alain.lepape@univ-tours.fr

Abstract

In this study we implemented an axillary SLN invasion model to develop highly sensitive imaging strategies enabling detection of a very small amount of tumor cells. A highly diffusible molecular probe targeting $\alpha v\beta 3$ and $\alpha v\beta 5$ integrins was investigated either via IV or locoregional injections. We additionally documented the potential interferences of this Near Infrared Fluorescence Probe with Blue Patente V and ICG dyes routinely used to facilitate lymph node detection during surgery. The human mammary adenocarcinoma MDA-MB-231-luc model was injected into the forepaw of nude female rats to obtain a controlled invasion of the axillary LN. Thanks to its high sensitivity, BLI was selected to achieve *in vivo* quantitation of tumor cells in SLNs and determine eligible animals for the study. NIRF of integrins was performed at 680 nm both *in vivo* and *ex vivo* using spectral unmixing to suppress auto-fluorescence signal and preserve sensitivity. *In vivo* BLI was quite reliable in estimating discrete invasion by cancer cells in the LN with thresholds of detection and quantitation of about 500 and 1500 cells respectively. For fluorescence at 680 nm, *in vivo* imaging is not suitable to detect micro-invasion, but *ex vivo* fluorescence with spectral unmixing of SLNs confirmed the presence of a tumor burden as low as 1500 cells expressing $\alpha v\beta 3/\alpha v\beta 5$ integrins. Targeting few tumor cells inside a micro-invaded sentinel lymph node by molecular probes is not sensitive enough to provide direct *in vivo* or peroperative imaging. At the time NIRF is performed on the excised specimen, high sensitivity imaging associated with spectral unmixing allowed such detection within less than 1 minute of examination.

---

*These authors contributed equally to this work.

1. Introduction

From the primary tumor, at the time tumor cells disseminate via lymph ducts, they accumulate in the first drainage site which is named the Sentinel Lymph Node (SLN). The interest in exploring the sentinel lymph node is to obtain information on the tumor invasion stage [1], because the invasion of SLNs by metastatic tumor is the major prognostic factor for patients with malignant tumors of epithelial origin [2]. In the clinical practice, the pathological examination of the excised specimen remains the gold standard for cancer confirmation. Apart from expensive fast molecular biology tests [3] [4], routine histopathology examinations require 4 - 7 days to detect minimal invasion and so does not allow the immediate assessment of the surgical outcome [5]. Positive micro-invasion of SLN in patients is relatively rare [6] [7] and the management of positive SLN in breast cancer has currently been revised. Most of the other SLN examinations are considered for gastric [8] [9], cervical, head and neck cancers [10] and melanoma [1]. Imaging strategies could provide information on the tumor invasion status faster than histopathology techniques. Nowadays, many different imaging agent sexist dedicated to the localization of LNs such as Blue Patente V (BPV) for visual direct observation or ICG (IndoCyanineGreen) for fluorescence peroperative imaging [11] [12]. For the imaging of invasion, labeled dedicated Monoclonal antibodies (MAbs) and RGD mimetic agents for targeting $\alpha V\beta 3$ and $\alpha V\beta 5$ integrins have been identified as efficient probes for *in vivo* imaging (PET, SPECT, MRI, NIRF). Targeting integrins from both tumor cells and tumor-related angiogenesis thank to low molecular weight highly diffusible probes should theoretically contribute to significantly increase the detection's sensitivity [13]. Among imaging modalities for such probes, Near Infrared Fluorescence Imaging (NIRF) is used more and more to investigate real-time peroperative visualization of tumors and SLNs [10] [14]. In this study, we performed preclinical investigations about assessment of tumor micro-invasion in a rat model of SLN. For such a purpose we considered Bioluminescence Imaging (BLI) thank to its very high sensitivity detecting a tiny number of tumor cells in order to select animals with a predefined SLN micro-invasion status. We then explored the ability of integrins NIRF to determine the invasion stage both *in vivo* and *ex vivo* on excised specimens. Considering the routinely used dyes such as Blue Patente V (BPV) by surgeons to evidence the SLNs, we investigated its potential quenching effect on fluorescence intensities from both an integrin targeting probe and ICG.

2. Materials and Methods

1) Ethics Statement

All procedures on animals were performed in accordance with European ethical guidelines (European directives 2010/63/EU) and were approved by the Regional Committee for Animal Care and Ethics in Animal Experiments (CECCO n˚ 3).

2) Cell Culture

The MDA-MB-231-luc-D3H2LN human breast adenocarcinoma cancer cell line was obtained from Perkin Elmer (France). The cancer cell line was maintained according to the supplier's instructions.

3) Animals

Pathogen-free 8 to 12 week-old female nude rats were purchased from Taconic (Germany). Rats were acclimated for 7 days in the laboratory before experimentation and were maintained in sterilized filter-stopped cages inside a controlled ventilated rack with access to food and water *ad libitum*. They were examined daily for clinical signs, distress, decreased physical activity and weighed 3 times a week.

4) Cell xenografts

Human breast cancer xenografts from MDA-MB-231-luc cells were established in NIH *nude* rats. We performed forepaw injection, so that 15 rats were anaesthetized by inhalation of 1.5% isoflurane with air (Isoflo®, AXIENCE S.A.S, France) and inoculated by 2×10^6 tumor cells in 10 μL PBS into the forepaw. The rats received pre-treatment with 10 mg/kg of busulfan 2 days before cell injection, so that immunodeficiency of rats was improved.

5) Bioluminescence Imaging

BLI was performed twice a week until the end of the study (Day 28) using an IVIS-Lumina II (Perkin Elmer, France). Each rat was IP injected with 20 mg luciferin potassium salt (Promega, France). Rats anesthetized by 1.5% isoflurane were placed on a thermostatically controlled heating pad (37˚C) during imaging. Acquisition binning and duration were set depending on tumor activity. Signal intensity was quantified as the total flux (photons/seconds) within ROIs drawn manually around the tumor area using Living Image 4.4 software (Perkin Elmer, France).

6) Estimation of the number of tumor cells present in the SLN

A calibration curve was determined from known amounts of cells (500, 1000, 5000, 10,000 & 50,000 cells) either deposited on the skin or subcutaneously injected. The injected volume was 100 μL of culture medium supplemented with luciferin potassium salt (300 μg/mL). Immediately after luciferin addition, the cell suspension was deposited or injected, and then BLI was performed 5 minutes later.

7) Near Infrared Fluorescence Imaging

NIRF was performed *in vivo* and *ex vivo* on excised specimens using an IVIS-Lumina II (Perkin Elmer, France) operated according to the spectral unmixing mode. Rats selected from BLI were IV injected with 20 nmol of IntegriSense680 (Perkin Elmer, France). Animals anesthetized by 1.5% isoflurane were placed on a

thermostatically controlled heating pad in supine or lateral position. Signal intensity was quantified as the total radiant efficiency ([photons/seconds]/[µW·cm^{-1}]) within ROIs drawn manually around LN area using Living Image 4.4 software (Perkin Elmer, France).

8) Assessment of fluorescence quenching

For these experiments, 50 µL of Blue Patente V (0.5 nmol, Guerbet, France), IntegriSense680 (0.5 nmol, Perkin Elmer, France), ICG (129 nmol, SERB France), or 50 µL of a mix of the two dyes (BPV + IntegriSense680 or BPV + ICG) were injected in the forepaw of the rat, and then axillary LNs were resected 15 min after injection, so that NIRF was performed on excised LNs.

9) Sacrifice and organ removal

Rats under anesthesia were sacrificed by lethal IV injection of pentobarbital (Ceva Santé Animale, Libourne, France) and LNs were collected from each animal for immediate *ex vivo* assessments.

10) Statistical analysis

Statistical analysis was performed using Graph Pad Prism software 5.0 (Graph Pad, USA). Statistical analysis was performed with the Student's unpaired t test.

3. Results

1) Assessments of the fluorescence quenching, and wash-out kinetics in lymph nodes from healthy rats:

There was no statistical difference in the fluorescence intensities quantified on *ex vivo* resected LNs after forepaw injection of IntegriSense680 versus BPV + IntegriSense680 (Figure 1(A)). There was no statistical difference in the fluorescence intensities quantified on *ex vivo* resected LNs after forepaw injection of ICG versus BPV + ICG (Figure 1(B)). Such an absence of quenching by BPV was also observed at the time a very low amount of IntegriSense680 was injected.

After forepaw injection of IntegriSense680, the dye diffuses mainly through the lymphatic ducts, then reaches LNs where it accumulates momentarily before being eliminated (Figure 2(A)). The comparison between NIRF signals in axillary and brachial LNs highlighted that the retention of the dye is greater in axillary LNs than in brachial LNs, furthermore there were no differences in signal intensities between 5H30 and 19H for these two LNs (Figure 2(B)). Following loco regional injection, the limited clearance in healthy LNs compromises a high sensitivity of detection of tumor cells in SLNs since the residual activity is about 1 log higher than predicted fluorescence from 1000 - 2000 tumor cells. So, only a systemic route such as IV resulting in quasi-negligible residual activity in a healthy LN can be considered.

2) Validation of BLI quantitation:

The minimal invasion status was determined thanks to *in vivo* calibration experiments by the quantification of known amounts of cancer cells suspended in a biological medium in the presence of luciferin. The bioluminescence signal was stable during at least 9 minutes following injection and attenuation of photons

Quenching assessment of IntegriSense680 Vs IntegriSense680+BPV
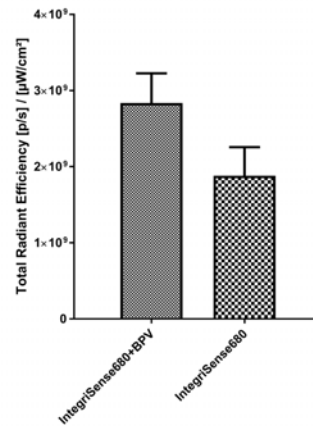
(A)

Quenching assessment of ICG Vs ICG+BPV
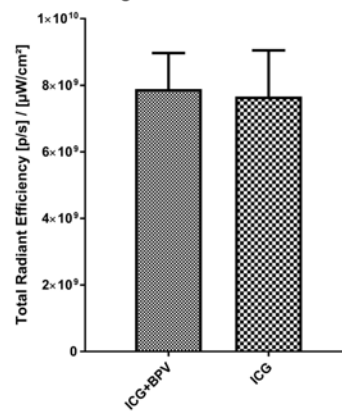
(B)

Quenching assessment of IntegriSense680 Vs IntegriSense680+BPV (2500fmol)
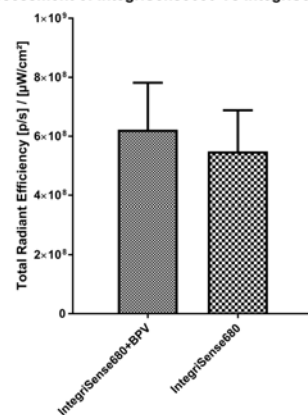
(C)

Figure 1. (A) Assessment of the fluorescence quenching of IntegriSense680 (0.5 nmol) by Patente Blue V (0.5 nmol). Quantifications performed by *ex vivo* NIRF on axillary LNs (n = 6 LN per group); (B) Assessment of the fluorescence quenching of free ICG (129 nmol) by Patente Blue V (0.5 nmol). Quantifications performed by *ex vivo* NIRF on axillary LNs (n = 6 LN per group); (C) Assessment of the fluorescence quenching of IntegriSense680 (injection of 2.5 pmol) by Patente Blue V. Quantifications performed by *ex vivo* NIRF on axillary LNs (n = 6 LN per group).
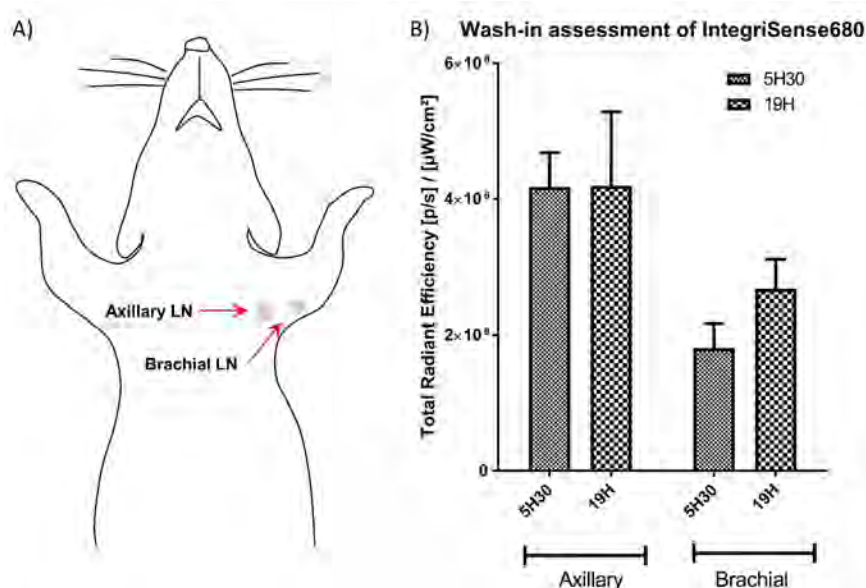
Figure 2. (A) Scheme showing the location of axial and brachial lymph nodes in the axillary area; (B) Comparison of the IntegriSense680 wash-in kinetics between axillary and brachial Lymph Nodes from rats by *ex vivo* NIRF quantification (n = 6 LN per group).

by skin was about 77% (**Figure 3(A)**). From the known number of SC injected cells, the calibration curve of bioluminescence activity corresponds to the following equation: $Y = 348X - 277556$ (**Figure 3(B)**) with a threshold of quantitation of about 500 cells (more than 5 times higher than background activity) (**Figure 3(C)**).

3) Sentinel Lymph Node invasion's model:

Thanks to BLI, it was possible to confirm the minimal invasion of tumor cells in both brachial and axillary LNs *in vivo* (**Figure 3(A)**). However, we noticed that tumor cells more often tended to develop in the axillary LN as compared to the brachial LN. From the *in vivo* BLI calibration curve, estimation of the tumor burden was approximately 18,700 cells for the rat A (**Figure 4(A)**) with a brachial invaded LN whereas rat B presented about 1500 cells (**Figure 4(B)**) in the axillary LN.

4) $\alpha V\beta 3$ and $\alpha V\beta 5$ integrins targeting:

Since locoregional injection of Integrin targeting probe IntegriSense680 results in a non-complete clearance in healthy LN, this route cannot be used to detect minimal invasion. So, only IV injections of the fluorescent probe were considered. In animals with SLNs containing almost 2000 cells, based on quantitative BLI, NIRF did not allow detection of any significant fluorescent signals *in vivo*. However, after excision of the SLN, *ex vivo* NIRF associated with spectral unmixing to suppress auto-fluorescence, enables confirmation of the minimal invasion of left axillary LN by cancer cells (**Figure 5**, ROI 2). Overall, the fluorescence activity of the minimally invaded LN ranged from 3 to 20 times higher than the control LNs. Fluorescence images from excised LNs were all in good correspondence with the *in vivo* bioluminescence patterns.
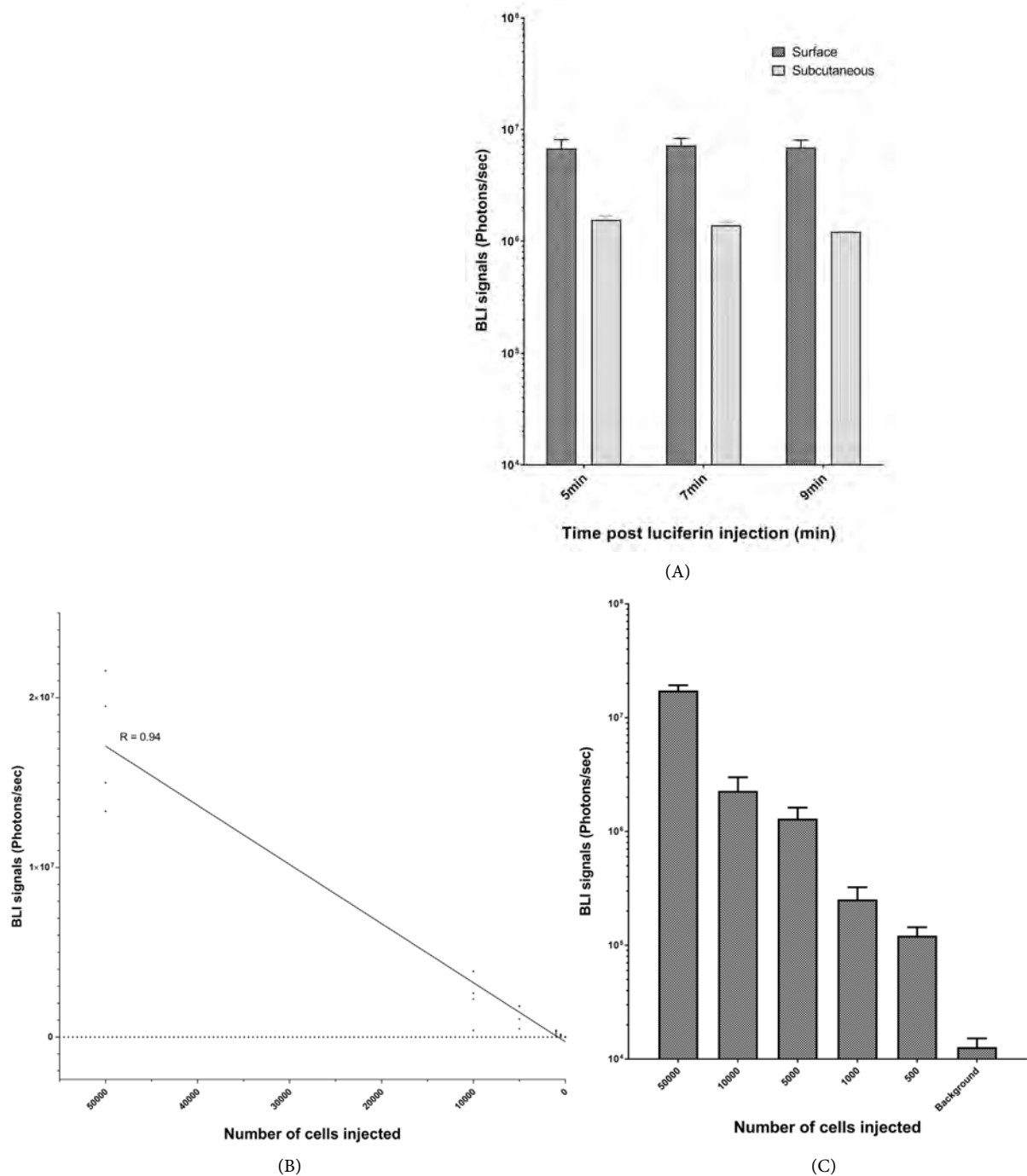
Figure 3. Validation of BLI quantification. (A) Comparison of *in vivo* BLI signals following subcutaneous injection or deposition of cells supplemented with substrate on the surface of the skin. (n = 2 per condition); (B) Determination of the fitting curve for *in vivo* quantitation of the number of cells (R = 0.94); (C) Assessment of BLI signal with different cell burdens following subcutaneous injection of cells supplemented with substrate. (n = 4 per condition).

4. Discussion

In clinical practice, blue dyes injected in the vicinity of primary tumors stain SLNs, so that clinicians can easily identify SLNs with the naked eyes during surgery [1]. When possible, radiolabeled colloids are co-injected allowing for the
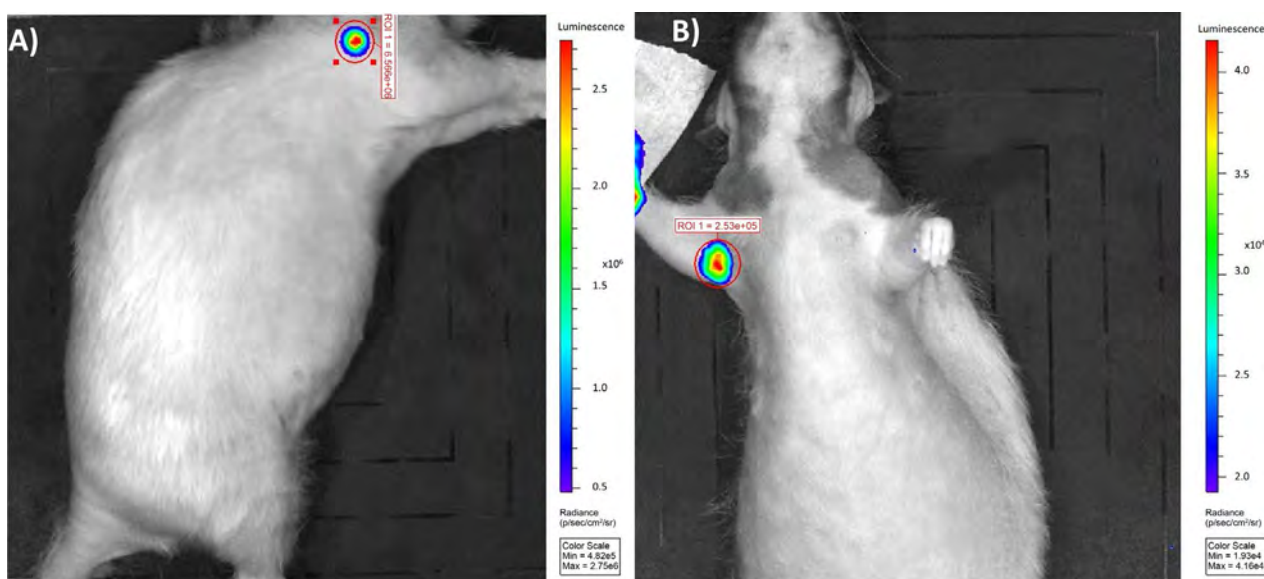
Figure 4. BLI of rats minimally invaded LN by cancer cells. To achieve more accurate quantitation, the best incidence for BLI acquisitions was anterior face for axillary LN and lateral for brachial LN. (A) BLI from right lateral side of rat A; (B) BLI from anterior side of rat B.
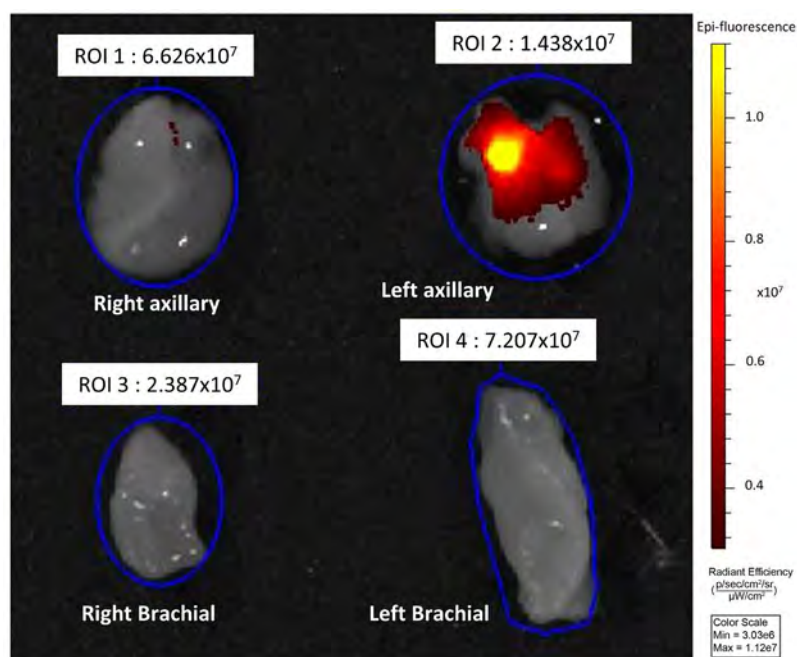
Figure 5. *Ex vivo* NIRF identification of tumor micro-invasion in the left axillary LN after IV injection of IntegriSense680 (20 nmol). Quantifications were performed following spectral unmixing.

detection of SLN location through the skin. Considering the technical and regulatory constraints associated with the use of radioisotopes, direct imaging of the SLNs via a fluorescent probe (*i.e.* ICG) should be a convenient and powerful alternative. To validate the use of infrared fluorescent probes at the same time as blue dyes, we assessed the potential quenching effect when either co-injected or

injected alone. Our results demonstrate the absence of quenching by BPV towards fluorescence of ICG and integrin targeting agents. So, fluorescence imaging is compatible with standard BPV-based surgery procedures.

When considering the simultaneous use of the two fluorescence probes ($ICG_{excitation} = 800$ nm; $ICG_{emission} = 843$ nm and $IntegriSense680_{excitation} = 675$; $IntegriSense680_{emission} = 693$ nm), optimized settings of excitation and emission filters allow for *in vivo* imaging of the SLN without reciprocal interferences. Due to the very lowIntegriSense680 fluorescence intensity present in micro-invaded SLNs as well as significant auto-fluorescence background, it is not possible to perform *in vivo* direct detection of IntegriSense680. So, *ex vivo* examinations on excised SLNs were performed using the spectral unmixing mode to suppress auto-fluorescence and contribution of ICG. In these conditions, the high quantum yield of fluorescence from IntegriSense680 is favorable for the level of detection requested and allows for evidencing of micro-invasion in SLNs as low as 1500 cells.

Fluorescent dyes with an emission of around 800 nm are usually more suitable for intraoperative imaging due to decreased auto-fluorescence of tissues and compatibility with lighting in the operating room [15]. However, higher wavelength targeting probes suffer from a decreased quantum yield (10 to 30 times lower) resulting in insufficient intrinsic sensitivity which is no longer suitable to detect a few tumor cells even on excised specimens. Obviously it is not possible to simultaneously consider the use of ICG and a specific targeting probe emitting fluorescence in the same wavelength window. NIRF for surgical applications is a modality with great potential, helping the surgeon's vision by providing real-time acquisitions with high contrasts. Molecular probes targeting tumors, especially fluorescent MAbs and RGD-derivatives would be promising to assess the tumor invasion stage [16].

5. Conclusion

BLI cannot be considered in humans but it is the most sensitive imaging modality to assess minimal tumor invasion (less than 1500 cells) for oncology research in animals. It was a valuable resource to acquire a relevant model with minimal invasion. NIRF of integrins targeted dyes in an invaded LN is possible *in vivo* or during surgery, but its sensitivity is not sufficient enough to demonstrate micro-invasion by a direct *in vivo* examination. When performed on an excised LN, *ex vivo* high sensitivity NIRF with spectral unmixing allows for detection of the micro-invasion.

Acknowledgements

Conflicts of Interest

None reported.

References

[1] Suman, B.M., Gao, S., Zhu, N., Liang, R., Gruev, V. and Achilefu, S. (2014) Real-Time Fluorescence Image-Guided Oncologic Surgery. *Advances in Cancer Research*, **124**, 171-211. https://doi.org/10.1016/B978-0-12-411638-2.00005-7

[2] David, S. (2003) Nathanson. Insights into the Mechanisms of Lymph Node Metastasis. *Cancer*, **98**, 413-423. https://doi.org/10.1002/cncr.11464

[3] Ohi, Y., Umekita, Y., Sagara, Y., Rai, Y., Yotsumoto, D., Matsukata, A., *et al.* (2012) Whole Sentinel Lymph Node Analysis by a Molecular Assay Predicts Axillary Node Status in Breast Cancer. *British Journal of Cancer*, **107**, 1239-1243. https://doi.org/10.1038/bjc.2012.387

[4] Brambilla, T., Fiamengo, B., Tinterri, C., Testori, A., Grassi, M.M., Sciarra, A., *et al.* (2015) One-Step Nucleic Acid Amplification in Breast Cancer Sentinel Lymph Node: A Single Institutional Experience and a Short Review. *Frontiers in Medicine* (*Lausanne*), **2**, 37. https://doi.org/10.3389/fmed.2015.00037

[5] Themelis, G., Harlaar, N.J., Kelder, W., Bart, J., Sarantopoulos, A., Van Dam, G., *et al.* (2011) Enhancing Surgical Vision by Using Real-Time Imaging of $\alpha v \beta 3$-Integrin Targeted Near-Infrared Fluorescent Agent. Translational Research. *Annals of Surgical Oncology*, **18**, 3506-3513. https://doi.org/10.1245/s10434-011-1664-9

[6] Lyman, G.H., Temin, S., Edge, S.N., Newman, L.A., Turner, R.R., Weaver, D.L., *et al.* (2014) Sentinel Lymph Node Biopsy for Patients with Early-Stage Breast Cancer. *Journal of Clinical Oncology*, **32**, 1365-1383. https://doi.org/10.1200/JCO.2013.54.1177

[7] Gojon, H., Fawunmi, D. and Valachis, A. (2014) Sentinel Lymph Node Biopsy in Patients with Microinvasive Breast Cancer: A Systematic Review and Meta-Analysis. *European Journal of Surgical Oncology*, **40**, 5-11. https://doi.org/10.1016/j.ejso.2013.10.020

[8] Takeuchi, H. and Yuko, K. (2015) Sentinel Lymph Node Biopsy in Gastric Cancer. *Cancer Journal*, **21**, 21-24.

[9] Fujimura, T., Fushida, S., Tsukada, T., Kinoshita, J., Oyama, K., Miyashita, T., *et al.* (2015) A New Stage of Sentinel Node Navigation Surgery in Early Gastric Cancer. *Gastric Cancer*, **18**, 210-217. https://doi.org/10.1007/s10120-014-0446-z

[10] Handgraaf, H.J., Boogerd, L.S., Verbeek, F.P., Tummers, Q.R., Hardwick, J.C., Baeten, C.I., *et al.* (2016) Intraoperative Fluorescence Imaging to Localize Tumors and Sentinel Lymph Nodes in Rectal Cancer. *Minimally Invasive Therapy* & *Allied Technologies*, **25**, 48-53. https://doi.org/10.3109/13645706.2015.1042389

[11] Niu, G. and Chen, X. (2015) Lymphatic Imaging: Focus on Imaging Probes. *Theranostics*, **5**, 686-697. https://doi.org/10.7150/thno.11862

[12] Li, L., Shiro Mori, S., Sakamoto, M., Takahashi, S. and Kodama, T. (2013) Mouse Model of Lymph Node Metastasis via Afferent Lymphatic Vessels for Development of Imaging Modalities. *PLoS ONE*, **8**, e55797. https://doi.org/10.1371/journal.pone.0055797

[13] Bach-Gansmo, T., Danielsson, R., Saracco, A., Wilsczek, B., Bogsrud, T., Fangberget, A., *et al.* (2007) Integrin Receptor Imaging of Breast Cancer: A Proof-of-Concept Study to Evaluate 99mTc-NC100692. *Journal of Nuclear Medicine*, **47**, 1434-1439.

[14] Atallah, I., Milet, C., Quatre, R., Henty, M., Reyt, E., Coll, J.L., *et al.* (2015) Role of Near-Infrared Fluorescence Imaging in the Resection of Metastatic Lymph Nodes in an Optimized Orthotopic Animal Model of HNSCC. E*uropean Annals of Otorhinolaryngology*, *Head and Neck Diseases*, **132**, 337-342.

https://doi.org/10.1016/j.anorl.2015.08.022

[15] Adams, K.E., Ke, S., Kwon, S., Liang, F., Lu, Y., Hirschi, K., *et al.* (2007) Comparison of Visible and Near-Infrared Wavelength-Excitable Fluorescent Dyes for Molecular Imaging of Cancer. *Journal of Biomedical Optics*, **12**, Article ID: 024017. https://doi.org/10.1117/1.2717137

[16] Desgrosellier, J.S. and Cheresh, D.A. (2010) Integrins in Cancer: Biological Implications and Therapeutic Opportunities. *Nature Reviews Cancer*, **10**, 9-22. https://doi.org/10.1038/nrc2748

Acute Mesenteric Ischemia: A Challenging Diagnostic Disease—Four Cases Reports and Literature Review (AMI)

Danilo Coco¹, Silvana Leanza²

¹Augusto Murri Hospital, Fermo, Italy
²Carlo Urbani Hospital, Jesi (Ancona), Italy
Email: webcostruction@msn.com, silvana.leanza@gmail.com

Abstract

Acute Mesenteric Ischemia (A.M.I.) is a potentially life-threatening condition syndrome due to inadequate or completely absent blood supply through superior or inferior mesenteric artery. The etiologies are various. Early diagnosis is essential to improve the clinical outcome. Despite advances in knowledge of pathophysiology, laboratory diagnosis and imaging techniques, acute intestinal ischemia is still associated with mortality rates.

Keywords

Acute Mesenteric Ischemia, Anatomy, Etiology, Treatment

1. Introduction

Acute Mesenteric Ischemia (A.M.I.) is a potentially life-threatening condition syndrome due to inadequate or completely absent blood supply through superior or inferior mesenteric artery, resulting in hypoperfusion and eventual ischemia/gangrene of the bowel. There are five types of A.M.I. subdivided for anatomic vessels or etiologies: Non occlusive Mesenteric Ischemia (N.O.M.I.) (20% - 30%), Acute Mesenteric Arterial Embolism (A.M.A.E.) (50%), Acute Mesenteric Arterial Thrombosis (A.M.A.T) (15% - 25%) and Mesenteric Venous Thrombosis (M.V.T.) (5%). The fifth type is only 3% and is a clinical entity that occurs as a consequence of mechanical scenarios such as strangulation, intussusception, volvulus, tumor or trauma. A.M.I. accounting for 0.1% of hospital admissions. In large part because of the association with atherosclerosis. AMI is commonly considered a disease of the older population, with the typical age of onset being older than 60 years. The symptoms are non-specific so the diagnosis

is very difficult .The radiologic image of choice is angio-TC-Scan. Treatment options include a range of options such as a few invasive management as percutaneous angioplasty plus stenting or thrombolytic therapy until surgical exploration [1].

2. We Demonstrate Four Clinical Scenarios to Underline the Difficult Way to Join a Diagnosis and Treatment

1) Clinical scenario

A 85 years old Caucasian male presented to the ED with significant periumbilical abdominal pain, vomiting and fever. He presented a medical history of atrial fibrillation and anticoagulant therapy. During the physical examination, the patient was uncomfortable and in severe sepsis. Her vital signs were: blood pressure, 80/77 mm Hg; respiratory rate, 30 breaths/minute; heart rate, 129 beats/minute; and temperature superior of 38°C. Oxygen saturation was 85% on room air. The abdominal examination revealed a distended abdomen that was severely tender to palpation, with rigidity, guarding, and rebound tenderness. Laboratory evaluation revealed high leukocytosis with a white blood cell (WBC) count of 30 per mm^3. Arterial Blood Gases (ABG) demonstrated metabolic acidosis. Computed tomography angiography (CTA) reveled a superior mesenteric artery completely occlused after aorta origin, dilatated ileum and abdominal fluid (Figure 1). The patient was immediately started intravenous (IV) fluids of 2 l in 6 hours, Foley and jugular cathether vein canulation to support main arterial pressure and urine output. The patient was transferred to surgical services. A xifo-pubic incision laparotomy was made. A totally ischemic small and partially large intestine was found. We made only a laparostomy for a second look. The patient died the day after.

2) Clinical scenario

A 74 years old woman was referred to our hospital because of sudden onset of continuous to severe abdominal pain. The pain was associated with nausea and passage of dark, loose stools. His medical history was significant for permanent

Figure 1. Superior Mesenteric Artery completely occluded after its origin from aorta.

atrial fibrillation (AF), Diabetes Mellitus type II and rheumatoid arthritis and an history of sigma resection plus colostomy and then reversal surgery after perforated diverticulitis. On physical examination an irregular rhythm of approximately 120 bpm was found. Electrocardiogram ECG revealed Atrial Fibrillation (AF). His arterial blood pressure was 140/85 mmHg. The abdomen was soft and mildly tender to deep palpation. Laboratory data revealed a raised white cell count (15 × 109/L, normal value 4.0 - 10.0 × 109/L), creatine kinase (300 UL, normal value < 130 UL) and lactate dehydrogenase (260 UL, normal value < 240). Arterial Blood Gas (ABG) was normally. Other laboratory parameters were in the normal range. The plain radiograph of the abdomen showed two air fluid levels. Abdominal ultrasound indicated no abnormalities. The patient underwent contrast enhanced abdominal CT. It showed superior mesenteric artery completely occluded after 6 cm of aorta, small intestine pneumomatosi and abdominal fluid (Figure 2(a) & Figure 2(b)). Treatment with subcutaneous analgesics, oxygen, intravenous broad-spectrum antibiotics and fluid resuscitation with crystalloids (50 - 70 mL/kg daily), according to hemodynamic monitoring,
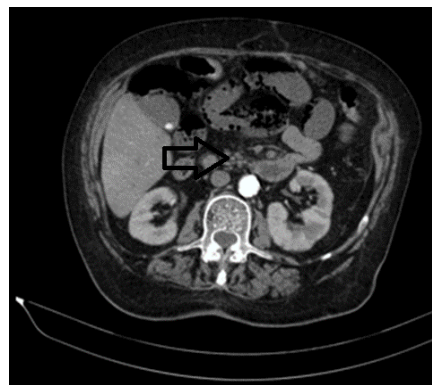
(a)
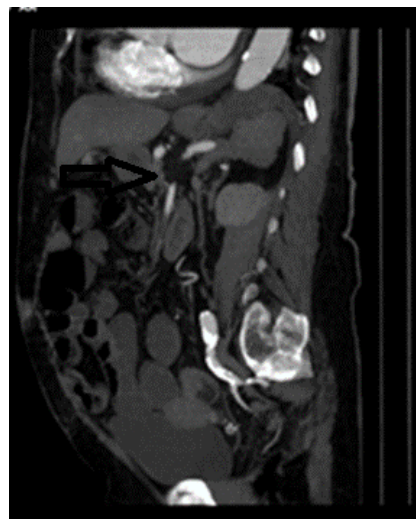
(b)

Figure 2. Inferior Mesenteric Artery (I.M.A.) not perfused after 6 cm at the origin of the aorta.

was commenced. A nasogastric tube was inserted to relieve distension. The patient was transferred to surgical services. After laparotomy, we found totally hypoperfused jejunum and ileum without necrosis. For this reason a laparostomy was made. 24 hours later, we performed a second-look surgery but we found a totally necrotic intestine.

3) Clinical scenario

A 63 years old man was admitted in our Emergency Room for acute abdominal pain associate with fever, vomiting, diarrhea and hypotension. His medical history reveled diabetes mellitus type II, hypertension, a history of Hairy Cell Leukemia with splenomegaly and cytopenia and endoprosthesis for thoracic aneurism. Physical examination demonstrated Temperature 37.5˚C, blood pressure 98/60 mmHg with mild orthostatic hypotension. Normal bowel sounds were noted in the abdomen, which was moderately tender in the right lower quadrant with focal rebound tenderness. The rectal exam was normal, with a stool sample negative for occult blood. Pelvic exam was unremarkable. WBC count was 24,300 per mm^3. Serum chemistries and arterial blood gases (ABG) were remarkable for a mild anion gap metabolic acidosis. Abdomen TC scan with artery phases was performed, revealing S.M.A. occluded at aorta origin, small bowel dilatation, pneumomatosi, peritoneal fluid (Figure 3(a) & Figure 3(b)). After resuscitation with crystalloid through jugular vein catheterization, we explored the abdomen. We found a 30 cm necrotic ileum and hypoperfusion wall in the other part of small intestine. We performed a 30 cm of small bowel resection and superior mesenteric artery longitudinal arteriotomy, Fogarty exploration with
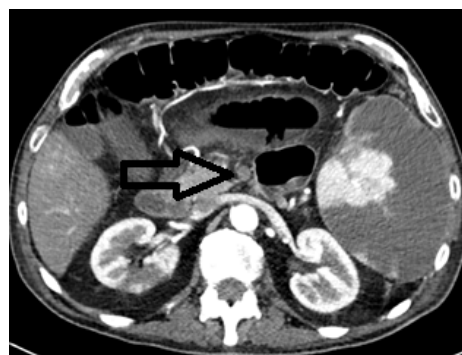
(a)
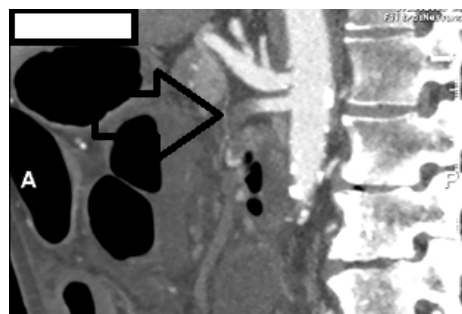
(b)

Figure 3. Superior Mesenteric Artery S.M.A. occluded.

thrombus extraction and a laparostomy. 24 hours later, we performed a second look finding an hypoperfusion small bowel. At third-look surgery we made resection of other 10 cm of small bowel and intestinal anastomosis. The patient was transferred in intensive care unit (ICU) where he died ten days later.

4) Clinical scenario

A 94 years old woman was referred to our hospital because of sudden onset of acute abdominal pain, dark stools and vomiting. She had a medical history of myocardial infarction. On physical examination, the abdomen was soft with moderate pain. Rectal exploration reveled mucorrea. Laboratory exams reveled only leukocytosis 11 per mm^3. ABG and abdomen plain was unremarkable. We performed an CT-Scan that demonstrated a distal occluded S.M.A. (Figure 4(a) & Figure 4(b)) Having considered the patient's age and in agreement with relatives, we treated her only with antithrombotic therapy and fluid resuscitation. She died one day later.

3. Anatomic Reasons for Intestinal Ischemia

Anatomy is crucial for the understanding of the pathophysiology, clinical presentation of intestinal ischemia. The splanchnic circulation is characterized by linking the 3 major vessels together [2]. The major connections are: the junction

(a)
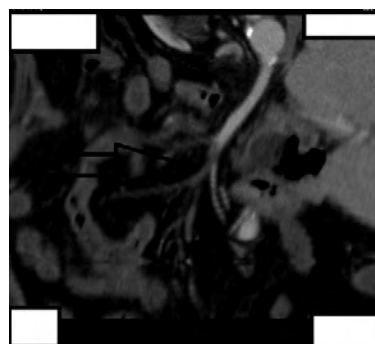
(b)

Figure 4. Superior Mesenteric Artery S.M.A. occluded distally.

of the superior and inferior pancreaticoduodenal arteries (linking the celiac axis and the superior mesenteric artery SMA), the marginal artery of Drummond and the Riolan's arch, between the middle colic and the left colic arteries (linking the superior mesenteric artery with the inferior mesenteric artery, IMA) and the collateralization between the inferior mesenteric artery and systemic circulation which occurs in the rectum (superior and middle rectal vessels) (Figure 5) [3].

4. Etiology

Based on its etiology, acute mesenteric ischemia can be caused by various type of pathology such as Atrial fibrillation/flutter, Myocardial infarction, Infective endocarditis, Mechanical valve prostheses, Cardiomyopathies, Valvular disease for Arterial embolism (A.M.A.E.) which represents 50% of cases. Atherosclerosis, Low cardiac output, Congestive heart failure, Procoagulative status, Vasculitis (including Takayasu arteritis), Aortic/superior mesenteric artery aneurysm/dissection, Cardiac catheterization and angiography for acute mesenteric thrombosis (A.M.A.T.) 15% - 25% cases. Cardiogenic shock, Hypovolemic shock, Septic shock, Cardiac/Major abdominal surgery, Dialysis, Vasoconstricting agents for N.O.M.I 20% - 25% of cases. Causes of Mesenteric venous thrombosis (M.V.T.) include Hypercoagulability, Tumor causing venous compression or hypercoagulability (paraneoplastic syndrome), Infection, usually intra-abdominal (e.g., appendicitis, diverticulitis, or abscess), Venous congestion from cirrhosis (portal hypertension), Venous trauma from accidents or surgery, Increased intra-abdominal pressure from pneumoperitoneum during laparoscopic surgery, Pancreatitis. The fifth type is only 3% and is a clinical entity that occurs as a consequence of mechanical scenarios such as strangling, intussusception, volvulus, tumor or trauma [4].
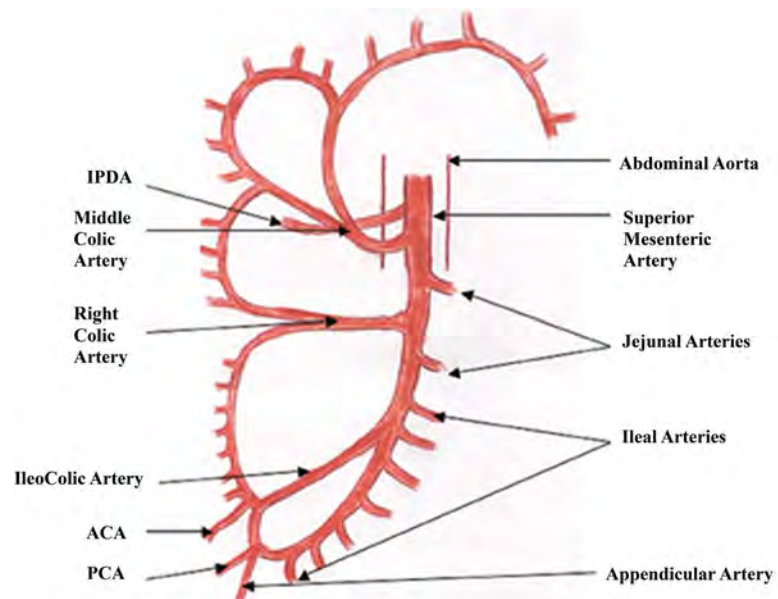

Figure 5. S.M.A. anatomy from Int. J. Morphol.
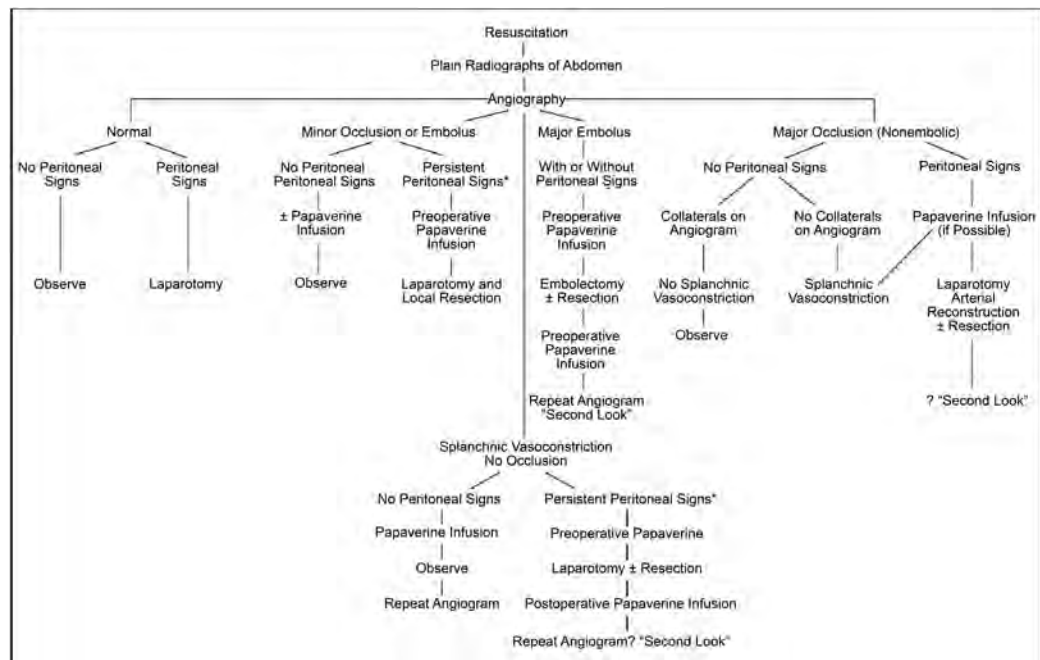
1) Clinical presentation

Symptoms and signs of mesenteric ischemia vary according to the etiology and the time of onset. They can be non diagnostic or non specific in almost 80% of cases. An abdominal pain was present in 95% (median of 24 hours duration). Other symptoms are nausea in 44%, vomiting in 35%, diarrhea in 35%, heart rate > 100 in 33%, "blood per rectum" and mucorrea in 16%, constipation in 7% [5].

2) Diagnosis

Laboratory studies are nonspecific. The most common laboratory abnormalities are hemoconcentration, leukocytosis, and metabolic acidosis, with high anion gap and lactate concentrations [6]. High levels of serum amylase, aspartate aminotransferase, lactate dehydrogenase, and creatine phosphokinase are frequently observed at presentation, but none is sufficiently sensitive or specific to be diagnostic In these four cases we found an altered Blood Test Count with WBC ranging from 10 to major 20 mmc in 90% of cases associated of increased lactic acidosis with arterial blood gases evidences metabolic acidosis but only in the advanced disease. In one cases a part of altered WBC the laboratory studies were normal. Normal laboratory values do not exclude this diagnosis and do not justify delaying angiography when clinical suspicion exists. MDCT (multidetector CT) scan angiography is first line diagnostic images to discover acute intestinal ischemia with sensitivity 83%, specificity 93%, positive predictive value (at prevalence of 62%) 93%, negative predictive value (at prevalence of 62%) 61%. Findings on CT scan include: Mesenteric edema, Bowel dilatation, Bowel wall thickening, Intramural gas, Mesenteric stranding, portal vein gas, abdominal fluid [7]. Angiography is a diagnostic and therapeutic method and it is considered the gold standard for the diagnosis of acute mesenteric ischemia associated with MDCT scan [8]. Its sensitivities in five of six studies have ranged between 90% and 100%; specificity was reported in two of these studies to be 100%. Although it has serious risks, angiography provides the possibility of direct infusion of vasodilators in the setting of non-occlusive ischemia [9].

4) Treatment

In the setting of suspected acute mesenteric ischemia, aggressive hemodynamic monitoring along with active resuscitation represents the first therapy. It should also include parenteral broad spectrum antibiotics to prevent bacterial translocation generated by ischemia associated with analgesia therapy. If there are no contraindications to anticoagulation, therapeutic intravenous heparin sodium should be administrated. The treatment of mesenteric ischemia depends on the causes and can be medical or surgical. In non-occlusive mesenteric ischemia, where there is no blockage of the arteries supplying the bowel, the treatment is medical. Thrombolytic medical treatment and vascular interventional radiological techniques have a growing role. When angiography is used to establish the diagnosis, the angiographic catheter should be left in the SMA for infusions of papaverine or other vasodilators. If the ischemia has progressed to the point that the affected intestinal segments are gangrenous, a bowel resection

Algorithm 1. Managment for the diagnosis and treatment of patients at risk of acute mesenteric ischemia (by "Acute Mesenteric Ischemia A Clinical Review" W. Andrew Oldenburg).

of those segments is called for [10]. Surgical revascularisation remains the treatment of choice for mesenteric ischaemia using longitudinal arteriotomy to perform embolectomy using a Fogarty catheter or distal bypass graft anastomosis. In most of cases multiple re-laparotomies are needed to obtain that news dead segments are removed after first operation and a second-look or even a third-look operation allows assessing segments that are borderline and that may be savable after revascularization [11].

5) Prognosis

Acute Mesenteric Ischemia (A.M.I.) is a potentially life-threatening condition syndrome. Prognosis is very poor and account for a mortality range from 50% to 90% for A.M.A.E, A.M.A.T and N.O.M.I and account of 30% for M.V.T. [12].

5. Discussion

Despite advances in knowledge of pathophysiology, laboratory diagnosis and imaging techniques, acute intestinal ischemia is still associated with mortality rates of more than 80%. Atrial fibrillation, myocardial disease or vascular disease and advanced age are the most common etiology. Differentiation between viable and nonviable, necrotic bowel tissue is done by exploratory laparotomy. A second-look laparotomy (after 24 - 48 hours) is recommended, even after successful primary intervention, because the intraoperative assessment of bowel viability [13]. Early diagnosis is the only best change but acute intestinal ischemia still remains a challenging diagnostic disease [14]. An algorithm management for the diagnosis and treatment of patients at risk of acute mesenteric ischemia is necessary to following the right steps (**Algorithm 1**) [15].

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Panes, J. and Pique, J.M. (2009) Intestinal Ischemia. In: Yamada, T., Ed., *Textbook of Gastroenterology*, 5 th Edition, Blackwell Publishing, Hoboken, 2811-2831. https://doi.org/10.1002/9781444303414.ch83

[2] Tendler, D.A. and Lamont J.T. (2009) Acute Mesenteric Ischemia. UpToDate. https://www.uptodate.com

[3] Rosenblum, J.D., Boyle, C.M. and Schwartz, L.B. (1997) The Mesenteric Circulation. Anatomy and Physiology. *Surgical Clinics of North America*, **77**, 289. https://doi.org/10.1016/S0039-6109(05)70549-1

[4] Boley, S.J., Brandt, L.J. and Veith, F.J. (1978) Ischemic Disorders of the Intestines. *Current Problems in Surgery*, **15**, 1-85. https://doi.org/10.1016/S0011-3840(78)80018-5

[5] Hunter, G.C. and Guernsey, J.M. (1988) Mesenteric Ischemia. *Medical Clinics of North America*, **72**, 1091-115. https://doi.org/10.1016/S0025-7125(16)30731-3

[6] Alpern, M., Glazer, G. and Francis, I. (1988) Ischemic or Infarcted Bowel: CT Findings. *Radiology*, **166**, 149-152. https://doi.org/10.1148/radiology.166.1.3336673

[7] Kao, L.S. and Lee, T. (2009) PreTest Surgery: PreTest Self-Assessment and Review. McGraw-Hill Medical, New York.

[8] Meng, X., Liu, L. and Jiang, H. (2010) Indications and Procedures for Second-Look Surgery in Acute Mesenteric Ischemia. *Surgery Today*, **40**, 700-705. https://doi.org/10.1007/s00595-009-4140-4

[9] Brandt, L.J. and Boley, S.J. (2000) AGA Technical Review on Intestinal Ischemia. American Gastrointestinal Association. *Gastroenterology*, **118**, 954-968. https://doi.org/10.1016/S0016-5085(00)70183-1

[10] Florian, A., Jurcut, R., Lupescu, I., Grasu, M., Croitoru, M. and Ginghină, C. (2010) Mesenteric Ischemia—A Complex Disease Requiring an Interdisciplinary Approach. A Review of the Current Literature. *Romanian Journal of Internal Medicine*, **48**, 207-222.

[11] Andrew Oldenburg, W., Louis Lau, L., Rodenberg, T.J., Edmonds, H.J. and Burger, C.D. (2004) Acute Mesenteric Ischemia A Clinical Review (REPRINTED). *Archives of Internal Medicine*, **164**, 1054-1062. https://doi.org/10.1001/archinte.164.10.1054

[12] Christensen, M.G., Lorentzen, J.E. and Schroeder, T.V. (1994) Revascularisation of Atherosclerotic Mesenteric Arteries: Experience in 90 Consecutive Patients. *European Journal of Vascular Surgery*, **8**, 297-302. https://doi.org/10.1016/S0950-821X(05)80145-8

[13] Endean, E.D., Barnes, S.L., Kwolek, C.J., Minion, D.J., Schwarcz, T.H. and Mentzer Jr., R.M. (2001) Surgical Management of Thrombotic Acute Intestinal Ischemia. *Annals of Surgery*, **233**, 801-808. https://doi.org/10.1097/00000658-200106000-00010

[14] Tola, M., Portoghese, A. and Maniga, A.M. (1997) Laparoscopic Second-Look in Acute Intestinal Ischemia [In Italian]. *Minerva Chirurgica*, **52**, 527-530.

[15] Seshadri, P.A., Poulin, E.C., Mamazza, J. and Schlachta, C.M. (1999) Simplified La-

paroscopic Approach to "Secondlook" Laparotomy: A Review. *Surgical Laparoscopy Endoscopy* & *Percutaneous Techniques*, **9**, 286-289.
https://doi.org/10.1097/00129689-199908000-00012

Fine-Grained Classification of Product Images Based on Convolutional Neural Networks

Tongtong Liu, Rubing Wang, Jikang Chen, Shengliang Han, Jimin Yang[*]

School of Physics and Electronics, Shandong Normal University, Jinan, China
Email: *jmyang@sdnu.edu.cn

Abstract

With the rapid development of the Internet of things and e-commerce, feature-based image retrieval and classification have become a serious challenge for shoppers searching websites for relevant product information. The last decade has witnessed great interest in research on content-based feature extraction techniques. Moreover, semantic attributes cannot fully express the rich image information. This paper designs and trains a deep convolutional neural network that the convolution kernel size and the order of network connection are based on the high efficiency of the filter capacity and coverage. To solve the problem of long training time and high resource share of deep convolutional neural network, this paper designed a shallow convolutional neural network to achieve the similar classification accuracy. The deep and shallow convolutional neural networks have data pre-processing, feature extraction and softmax classification. To evaluate the classification performance of the network, experiments were conducted using a public database Caltech256 and a homemade product image database containing 15 species of garment and 5 species of shoes on a total of 20,000 color images from shopping websites. Compared with the classification accuracy of combining content-based feature extraction techniques with traditional support vector machine techniques from 76.3% to 86.2%, the deep convolutional neural network obtains an impressive state-of-the-art classification accuracy of 92.1%, and the shallow convolutional neural network reached a classification accuracy of 90.6%. Moreover, the proposed convolutional neural networks can be integrated and implemented in other colour image database.

Keywords

Product Classification, Feature Extraction, Convolutional Neural Network (CNN), Softmax

1. Introduction

With the popularity of the Internet and varieties of terminal equipment, online shopping has become a regular part of people's lives with the onset of websites such as Amazon, Dangdang, Taobao, and Jingdong. Customers view a large number of product images, and there is an urgent need for efficient product image classification methods. At present, most studies have mainly focused on keyword-based, label-based, and content-based image retrieval. Zhou [1] used a querying and relevance feedback scheme based on keywords and low-level visual content, incorporating keyword similarities. He [2] proposed a method based on the Multi-Modal Semantic Association Rule (MMSAR) to automatically combine keywords with visual features automatically for image retrieval. Xu [3] used Bayes with expectation maximization to learn an initial query concept based on the labeled and formerly unlabeled images, and the active learning algorithm selects the most useful images in the database to query the user for labeling. However, the keywords and labeled information can only explain the basic information of the goods, such as the name of the product name, the origin, the size, and price and so on, which are difficult to reflect the complete characteristics of the products. At last, images have more information and intuitive expression. If we set an image classification filter on a shopping website, it will be convenient for users to browse and quickly find their favorite products.

The last decade has also witnessed great interest in research on content-based image classification. Image classification based on the content is based on the image features, including image shape, color, and texture.

Jia [4] adopted a gist descriptor and three complementary features, including Pyramid Histogram of Orientated Gradients (PHOG), Pyramid Histogram of Words (PHOW), and Local Binary Pattern (LBP) to extract and describe the features of product images. Valuable product information (such as long skirts versus skirts, and turtleneck versus round collars) can be labeled based on the image features and classification algorithms. Furthermore, they combine discriminative features for the SVM classifier. Experimental results showed that the performance of the product image database (PI 100) improved significantly using features fusion.

Nilsback and Zisserman [5] used the features of the Histogram of Gradient Orientations (HOG), HSV value, and Scale Invariant Feature Transform (SIFT) combining an SVM classifier and multiple-kernel learning framework to classify flower images. The classification accuracy ranged from 76.3% to 95.2%.

Yao and Khosla [6] proposed a random forest, in which every tree node is a discriminative classifier that can combine node information and all upstream nodes. This method identified meaningful visual information of both subordinate categorization and the activity recognition database.

For fine-grained classification, Yao [7] presented a codebook-free and annotation-free approach for fine-grained image categorization of birds. Experimental results showed that the method was better than state-of-the-art classification

approaches on the Caltech-UCSD Birds database. Krause and Stark completed the fine-grained classification of 3D cars [8].

Dyrmann and Karstoft [9] presented a method that recognized a large number of plant species in color images, by designing and training a deep convolution neural network. The network achieved a classification accuracy of 86.2% for a total of 10,413 images containing 22 species.

However, it was difficult to contain features like the above-mentioned shape, color, and texture that could be applied to all product image classifications. Compared to these classification methods, Convolutional neural networks (CNNs) are one of the deep learning algorithms with strong ability to acquire features, simple structure, and few parameters [10]. Nevertheless the fine-grained classification of a category in product images is rarely observed.

In recent years, CNNs received much attention on the computer vision research community, mainly because they have proven to be capable of effectively classifying images and outperforming previous records in image recognition challenges. Most noticeably is the task by Krizhevsky, Sutskever, and Hinton [11], who in 2012 had a margin of 10.9% compared to the second-best entry in the ImageNet Large Scale Visual Recognition Challenge [12]. The ImageNet challenge distinguishes objects such as cat, car, tree, and house from 1000 different categories. CNNs are a deep learning application to images, and they stimulate the neuron's activity in the neocortex, where most thinking happens, as Lecun describes [13]. The main benefit of using CNNs is that they are traditional, fully connected neural networks and can reduce the amount of parameters to be learned. Convolution layers effectively extract high-level features with small-sized kernels and feed the features to fully connected layers. According to Rumelhart, Hinton, and Williams [14], the training of CNNs is performed through back-propagation and stochastic gradient descent.

This study proposed a novel deep CNN that has data augmentation pre-processing, feature extraction, and softmax classification. To solve the problem of long training time and high resource share of deep convolutional neural network, this paper designed a shallow convolutional neural network to achieve the similar classification accuracy. To evaluate the classification performance of the networks, experiments were conducted using a public database Caltech256 and a homemade product image database from shopping websites.

2. Data Material

2.1. Caltech256 Database

The Caltech256 database containing 256 object categories on a total of 30607 images. This paper selected 20 object categories which were similar to product images to input the deep convolutional neural network for training. Each category of images was randomly selected 100 images for training and 50 images for testing. The size of the input image was normalized to 256 × 256 during the experiment.

2.2. Homemade Database

The data used in the numerical analysis are mainly obtained from Internet-based e-commerce databases, including T-mall, Jingdong, and Amazon. As shown in Figure 1, 20 products were selected, including garments and shoes. The garments consist of trousers, sweaters, jackets, outdoor jackets, dresses, short T-shirts, down jackets, fleeces, vests, Chinese dresses, shirts, short pants, short skirts, scarves, and socks. The shoes include skateboard shoes, basketball shoes, leather shoes, climbing shoes, and running shoes. Each product has 200 images and after using data augmentation, there were 20,000 product images in which 16,000 images were used for training purposes and 4000 images for testing purposes. The image sizes are not the same. To facilitate the experiment, all images are normalized into $256 \times 256 = 65,536$ pixels.
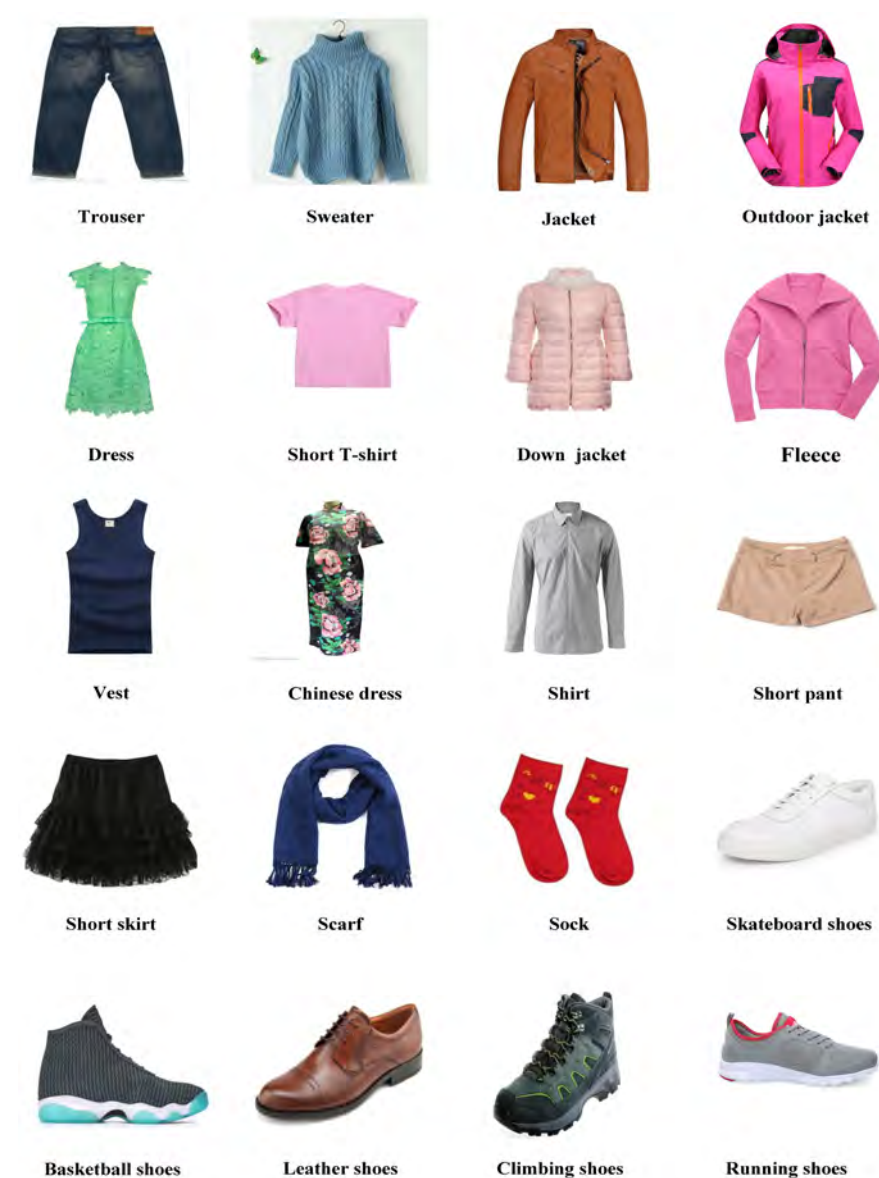

Figure 1. Garments and shoes database.

3. Methods

This section describes the pre-processing of the product images and the architecture of the deep convolutional neural network is used for classification of the garments and shoes.

3.1. Data Augmentation

A CNN is translation invariant but not rotation invariant. The number of product images, however, on the website is limited and therefore, we can generate the training and testing data by rotating the original data using affine transformation. The data was thereby increased five-fold by mirroring the images horizontally and vertically and rotating them in 90° and 180° increments. After using data augmentation, there were 20,000 product images in which 16,000 images were used for training and 4000 images for testing.

3.2. Model Architecture

Several pre-trained networks for image classification exist such as AlexNet [11], VGGNet [15] and GoogleNet [16], which won the championship in the ImageNet Image Recognition Competition in different years. In 2012, the Hinton Task Force participated in the ImageNet Image Recognition Competition for the first time to demonstrate the potential of deep learning. It won the championship by building the CNN network AlexNet, which consisted of 8 layers and reached a 16.4% error rate. The basic composition of VGGNet is similar to AlexNet, and is also characterized by continuous convolution and large amounts of computation. VGGNet consists of 19 layers and reached a 7.3% error rate. GoogleNet, the champion model of the ImageNet competition in 2014, proved that more convolution and deeper levels can obtain a better structure. Table 1 presents the performance results of AlexNet, VGGNet, and Google Net in the ImageNet image Recognition Competition.

However, the pre-trained network was created by the ImageNet, which is the largest database of image recognition in the world, which is different from the images in this study. Therefore, a new architecture was built to create a better classification of product images. Our CNN is sketched in Figure 2.

The images in the database are 256 × 256 RGB images. Matlab is used to augment data and transform the data into 227 × 227 RGB images. The network accepted 227 × 227 RGB images as input and output a vector for each block, as illustrated in Figure 2. The network had one 7 × 7 convolution layer with a stride

Table 1. Performance of Alex NET, VGG Net and Google Net.

	Alex Net	VGG Net	Google Net
Number of layers (layer)	8	19	22
Filter size	11,5,3	3	7,1,3,5
Top-5 Error rate (%)	16.4	7.3	6.7
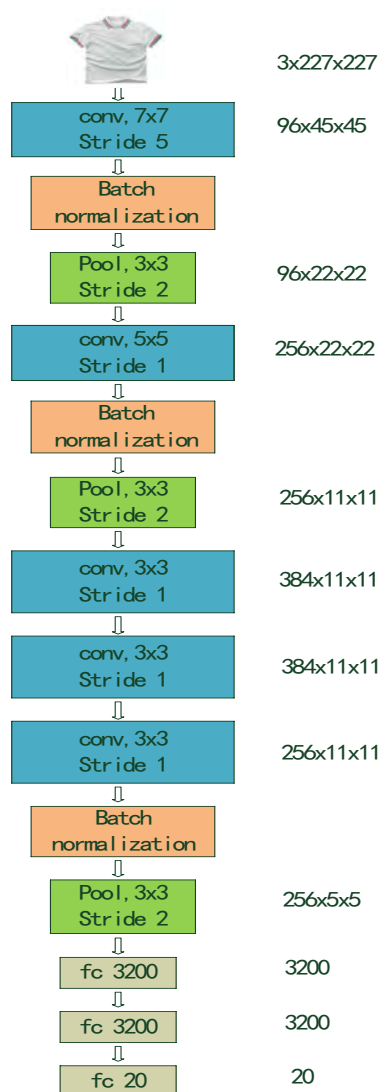
Figure 2. Deep convolutional neural network architecture.

of 5, followed by a 3 × 3 max-pooling layer with a stride of 2. This was mapped into a 5 × 5 convolution layer, which increased the number of filters from 96 to 256. Next, a 2 × 2 max-pooling layer was mapped with a stride of 2, and the number of filters was increased from 256 to 384. Following this, there were three 3 × 3 convolution layer with strides of 2. Finally, the network contained three fully connected layers, consisting of 3200 neurons and 20 softmax classifiers. In total, the network contained 628,324 learnable parameters, which is small compared to the 60 M parameters of AlexNet.

During the training, a 50% dropout was used before the three fully connected layers. The hidden layer was randomly discarded with 50% for training every epoch. This prevents all feature selectors from amplifying or reducing features all of the time, which are over fitted in the case of small samples and made poor generalizations. The dropout was used to avoid these problems in the training process. The network was trained using mini-batches with 128 images per batch,

to speed up the gradient update with a learning rate set to 0.001.

3.3. Input Layer

Data is fed to the network and the input layer produces an output vector as input to the convolution layer. Input data can be either raw image pixels or their transformations, which emphasize specific aspects of the image. This study inputs three-channel product images through a data augmentation method.

3.4. Convolution Layers

The convolution layer is the feature extraction layer. The input of each neuron is connected to the local receptive field of the previous layer, and the local feature is extracted. One of the important features of the convolution operation is that it enhances the original signal characteristics and reduces the noise. Filter kernels are slid over the original image and for each position, the dot product between the filter kernel and the part of the image covered by the kernel is determined. The calculation of the convolution layer is

$$x_j^{(l)} = f\left(\sum_{i \in M^{l-1}} x_i^{l-1} * k_{ij}^{(l)} + b_j^{(l)} \right) \tag{1}$$

where l is the number of layers, k_{ij} represents a convolution kernel with the connection of map j in the l layer and map i in the $l-1$ layer, x^{l-1} is the input feature maps of the $l-1$ layer, * represents convolution, b is the bias, and $f(\cdot)$ is the nonlinear activation function.

3.5. Max-Pooling Layers

The max-pooling layer is a method of aggregate statistics that uses the maximum or mean value of the region to reduce spatial size of a feature map and provide invariance to the network. Max-pooling layers can reduce the image size of the next layer, thereby reducing the parameters and calculations of the network. This is done by only keeping the maximum value within a $k \times k$ neighborhood in the feature map.

3.6. Batch Normalization

The role of batch normalization [17] is to normalize input data in the same range, even though the earlier layers were updated. According to Dieleman S, during each stochastic gradient descent (SGD), the corresponding activation was normalized by the mini-batch, so that the mean value of the result (output signal in each dimension) was 0 and the variance was 1 [18]. The calculation of the batch normalization is

$$y = \frac{\gamma}{\sqrt{\sigma^2 + \varepsilon}} x + \left(\beta - \frac{\gamma \cdot \mu}{\sqrt{\sigma^2 + \varepsilon}} \right) \tag{2}$$

where μ and σ are the mean value and variance of the image batch x, and γ and β are trainable parameters that are updated after each batch. ε is a small constant

value that is added to the variance to avoid division by zero.

3.7. Activation Functions

The activation functions in deep learning are responsible for applying a non-linear function to the output of the previous layer. Sigmoid, tanhyperbolic (tanh), rectified linear unit (ReLU) and softplus are commonly used in deep learning.

The non-linear Sigmoid function has a large signal gain in the central region, and relatively small signal gain on both sides [19]. The output of the sigmoid function is mapped into the internal of 0 and 1, so it has a good effect on the feature space map of the signal. However, this kind of activation function cannot solve the vanishing gradient problem and is slow in network training. The calculation of the sigmoid function is

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

The non-linear tanh function converges faster than the sigmoid function. It is mapped into the internal of −1 and 1, and the output is centered at 0. Still, the tanh function (like the sigmoid function) cannot solve the vanishing gradient problem. The calculation of the tanh function is

$$\tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{4}$$

This study used ReLU as the activation function. In 2011, the ReLU activation function was proposed by Glorot [20]. According to Krizhevsky [11], the ReLU function effectively suppressed the vanishing gradient problem with a faster convergence rate in training gradient descent than traditional saturated nonlinear functions. They can speed up training and keep the gradient relatively constant in all network layers. The ReLU is defined as

$$f(x) = \max(0, x) \tag{5}$$

The rectifier function is one-sided and therefore does not enforce a sign symmetry or antisymmetry. However, the response to the opposite of an excitatory input pattern is 0 (no response). Therefore, it is more biologically plausible and provides good results.

A smooth approximation to the rectifier is the softplus function. The softplus is not completely one-sided, so it is less biologically plausible and is not used as widely as ReLU. The calculation of the softplus function is

$$\text{softplus}(x) = \log(1 + e^{x}) \tag{6}$$

where x is the value of input signal.

Figure 3 shows the corresponding curves of the activation functions.

3.8. Fully Connected Layers

According to traditional neutral networks, all inputs in fully connected layers are
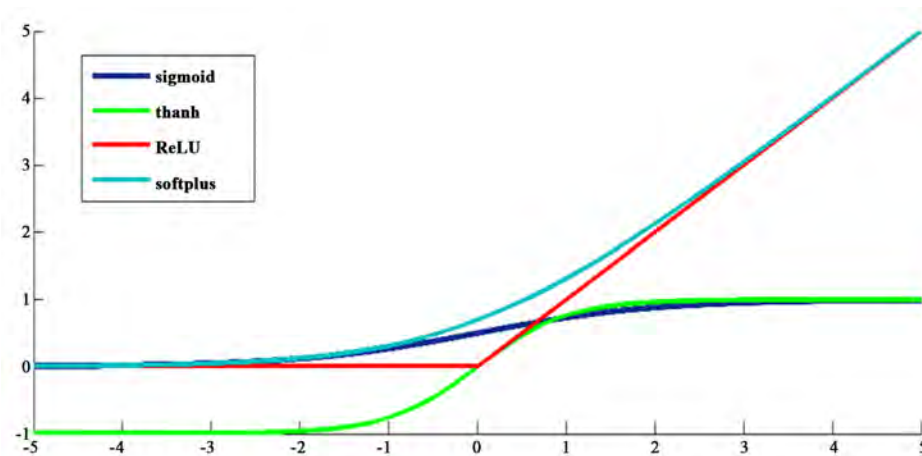
Figure 3. Activation function curves.

connected to all outputs of the previous layer. The fully connected layers are used as a way of mapping spatial features to image labels. After being trained, the network can extract features in these layers to train another classifier.

3.9. Softmax

This study used the softmax classifier, which is the generalization of the logistic model on multiple classification. The softmax classifier is an algorithm that divides the target variable into several classes. Supposing there are N input images $\left\{x_i, y_i\right\}_{i=1}^{N}$, each image is marked with k classes $y_i \in \left\{1, 2, 3, \cdots, k\right\}, k \geq 2$; in this study, $k = 2$. For the given test image x_i, the approximate value $p\left(y_i = j \mid x_i\right)$ of each class *j* is estimated by the hypothetical function. The calculation of the hypothetical function $h_\theta\left(x_i\right)$ is

$$h_\theta\left(x_i\right) = \begin{bmatrix} p\left(y_i = 1\right)\big|x_i;\theta \\ p\left(y_i = 2\right)\big|x_i;\theta \\ \vdots \\ p\left(y_i = k\right)\big|x_i;\theta \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x_i}} \begin{bmatrix} e^{\theta_1^T x_i} \\ e^{\theta_2^T x_i} \\ \vdots \\ e^{\theta_k^T x_i} \end{bmatrix} \tag{7}$$

where $\dfrac{1}{\sum_{j=1}^{k} e^{\theta_j^T x_i}}$ represents the normalization of the probability distribution,

that is, the sum of all probabilities is 1. θ is a parameter of the softmax function. The calculation of the loss function is

$$J\left(x, y, \theta\right) = -\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{j=1}^{k} 1\left\{y_i = j\right\} \log \frac{e^{\theta_j^T x_i}}{\sum_{j=1}^{k} e^{\theta_j^T x_i}}\right] \tag{8}$$

where $1\left\{y_i = j\right\}$ is an indicative function. The rule of value is as follows: $1\left\{\text{the value of expression is true}\right\} = 1$, $1\left\{\text{the value of expression is false}\right\} = 0$. Finally, the error function is minimized by stochastic gradient descent.

3.10. Filter Capacity

In this study, the efficiency of the network was determined by evaluating the filter capacity and coverage of the network [21]. The filter capacity is a measure of the filter's ability to detect complex structures in an image. If the capacity is small, only local features in the image will be mapped to the next layer. On the contrary, if the capacity is large, the filter will find complex structures of elements that are not neighbors in the input image. The filter capacity is calculated as the ratio between the real filter size and the receptive field [22]. The calculation of the capacity is

$$Capacity = \frac{real\ filter\ size}{receptive\ field} \tag{9}$$

where the real filter size is the size of the kernel, which consists of downsampling (striding or pooling) of previous layers. If no downsampling is applied, the real filter size is the same as the kernel size. For example, if the input to a layer with kernel size $n \times n$ is downsampled by a factor k, the real filter size would then be $kn \times kn$. In this network, there are two 3×3 max-pooling layers and a 2×2 max-pooling layer. After the first 3×3 max-pooling layer, the real filter size would be $3n \times 3n$. After the second 2×2 max-pooling layer, it would be $6n \times 6n$ and after the third 3×3 max-pooling layer, it would be $18n \times 18n$. The receptive field is defined as the region in the original image that a particular CNN's feature is focused on [22]. Increasing the size of filters in the convolution layers or using pooling can increase the receptive field and thus the filter capacity. According to Cao [21], the network is meaningless if the capacity is smaller than 1/6. For this network, the filter capacity is between 20.4% and 100%, and thereby well above the lower 1/6 limit.

3.11. Coverage

Coverage is a measure to "see" a part of the input image of the layer in a CNN. Adding convolution or pooling layers can increase coverage. The coverage of the network in the end should not exceed 100%. If coverage exceeds 100%, it will be a waste of network calculations, because the network can operate images larger than the input image. For this network, the convolution filters covered 55.9% of the input image and never exceeded the size of the image. Table 2 shows the coverage and capacity of the network.

4. Results and Discussion

The operating system used in the experiments is Centos 7, and four Navida TITIAN X graphics cards are used. The framework used is caffe, and the analysis of the experimental results is all based on caffe.

The classification accuracy of deep convolutional neural network on Caltech256 database reached 94.8%. It shows the effectiveness of the proposed deep network and its suitability for feature extraction of color images.

Table 2. Coverage and Capacity of the Network.

	Coverage (%)	Capacity (%)
Conv1	3.08	100
Conv2	25.1	26.3
Conv3	42.7	27.8
Conv4	51.5	23.1
Conv5	60.4	19.7

Figure 4 shows the classification accuracy and cross entropy loss of the experiment on homemade database. To achieve the highest accuracy possible without overfitting the network, the training was set to 100 epochs. The average classification accuracy of the test was 92.1%. Setting appropriate learning rates in the experiment can improve the learning efficiency of network and therefore improve the classification accuracy. The learning rate was reduced three times before the experiment was stopped. At the beginning, we set the learning rate at 0.001. The test accuracy rapidly increased and the test loss rapidly declined. According to the decline of train loss curve, the learning rate of the network is relatively high. After 10 epochs, the test accuracy slowly increased, even decreasing, and the test loss was an upward trend. We therefore set the learning rate at 0.0005. It was observed that the test accuracy of the network increased again and the test loss slowly decreased. After 20 epochs, the test accuracy and test loss was not stable. We set the learning rate at 0.0001. It was observed that the test accuracy was high and the train loss continued to decline, then stabilized after 30 epochs.

Figure 5 shows the confusion matrix of the misclassification fraction for each of the 20 species. Here, it is seen that jackets (#2), shirts (#10), short pants (#11), short skirts (#12), and socks (#14) were often correctly classified with an accuracy of 96%, 96%, 97%, and 95%, respectively. However, there was no clear species that trousers (#0), sweaters (#1), outdoor jackets (#3), dresses (#4), short T-shirts (#5), down jackets (#6), fleeces (#7), vests (#8), and scarves (#13) was confused with. The classification for these species ranged from 90% to 95%. Chinese dresses (#9), skateboard shoes (#15), basketball shoes (#16), and leather shoes (#17) were often misclassified. Of these three species, only 89%, 85%, 82.5%, and 88.5% were classified correctly. Skateboard shoes (#15) were often classified with leather shoes (#17), and basketball shoes (#16) were often classified with skateboard shoes (#15), because they are similar in shape and texture. Leather shoes (#17) were often misclassified with climbing shoes (#18) and running shoes (#19), because they are similar in color and shape. The classification accuracies for these three species were, however, still well above random assignment.

Overall, most species had the highest classification accuracies. This is because the aim of the training was to obtain the most correctly classified product images, without taking into account how these product are distributed among the
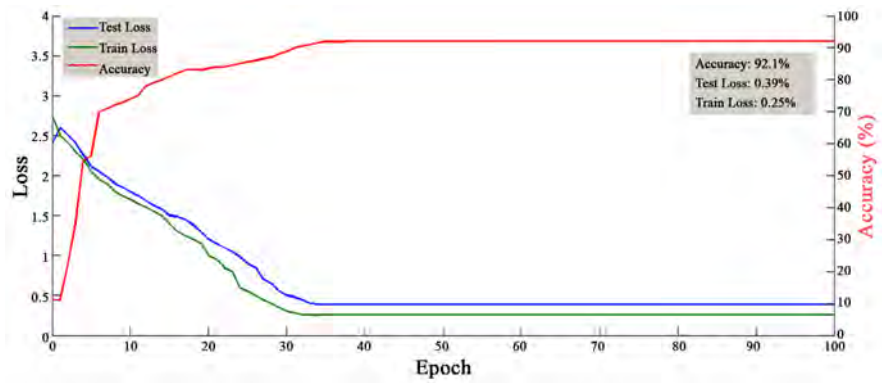
Figure 4. Classification accuracy and cross entropy loss of experiment. Red line represents the test accuracy, blue line represents test loss, green line represents train loss.
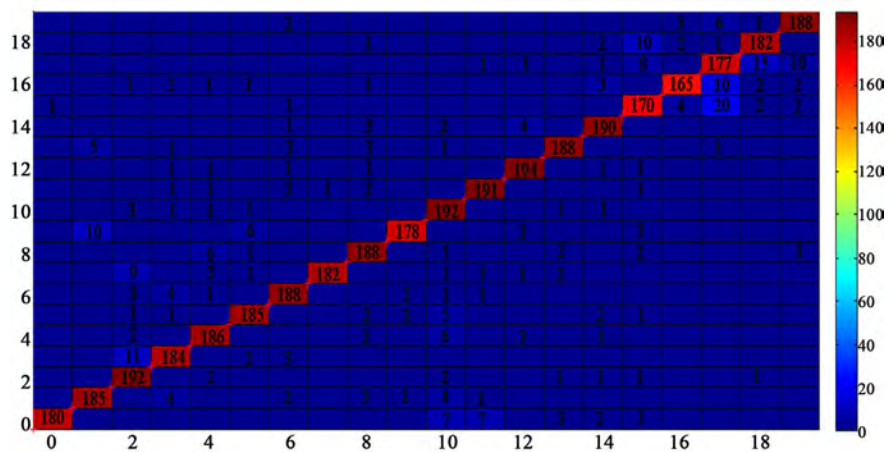

Figure 5. Confusion matrix.

20 classes. Few image samples contributed less to the overall loss. The average classification accuracy of product images was 92.1%. The classification accuracy of garments was 93.4%, and the accuracy of shoes was lower at 88.2%. This was because the shoes sample that we chose was similar and the features could not be better extracted.

As shown in **Figure 6**, we chose an image from each of the three categories, including short skirt, trouser, and basketball shoes, to show the visualization feature images of each convolution layer. It can be seen from the horizontal comparison of the feature images of each category that the first convolution layer (conv1) shows the edges, shapes, and colors of the product. Conv2 shows the texture of the product. After conv3, the feature images of product are more ambiguous and have no specific meaning. The classification accuracy of short skirts, trousers, and basketball shoes was 97%, 90%, and 82.5%. It can be seen from the vertical comparison of the feature images of each category that the edge sharpness of skirt is higher than trouser, and the trouser is higher than basketball shoes after conv3. It can be also proven from **Table 3**, which shows the mean and standard deviation value of each convolution layer feature extraction.
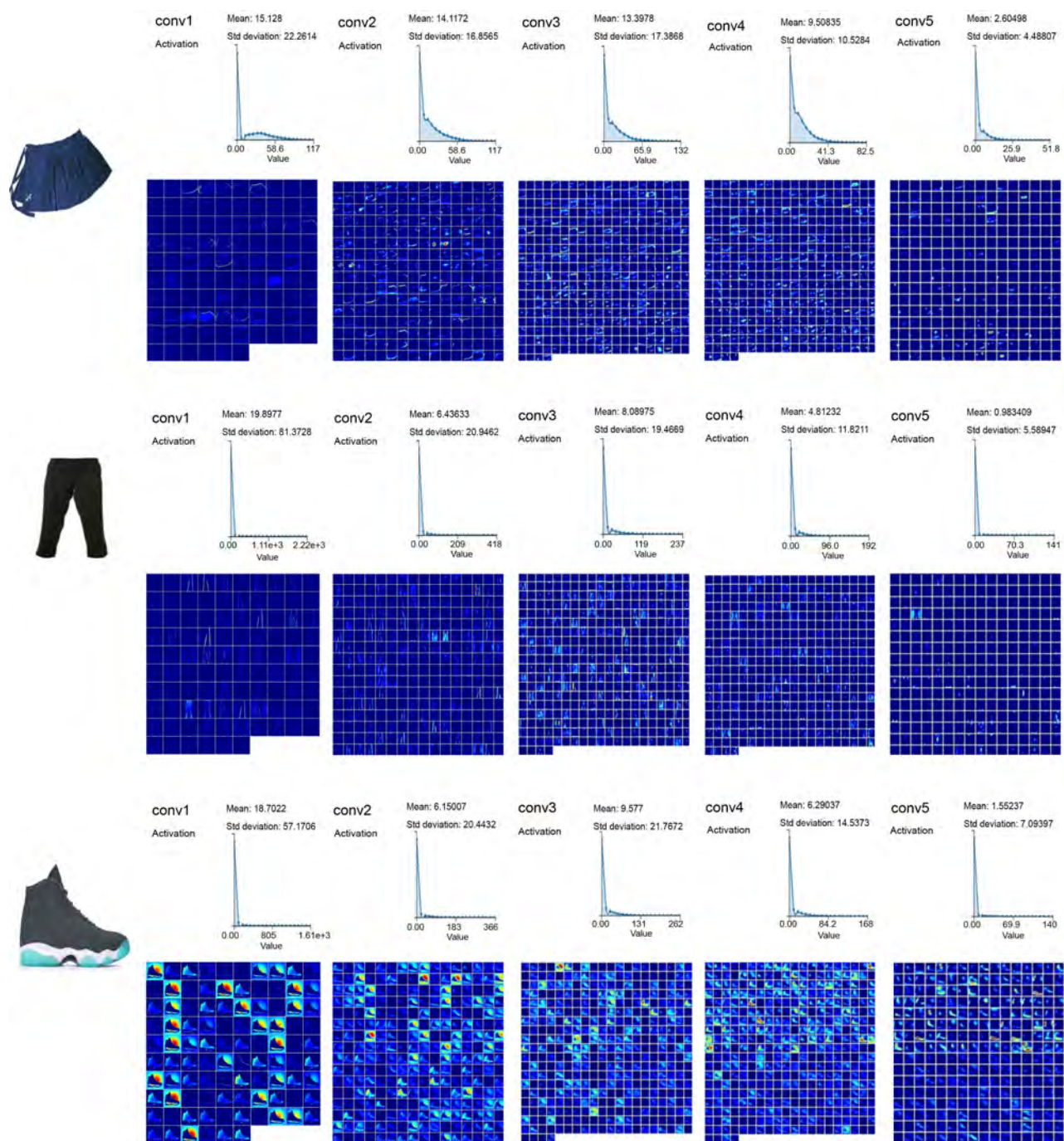
Figure 6. Visualization feature images of each convolution layer. (a) Short skirt; (b) Trouser; (c) Basketball shoes.

Table 3. The mean and standard deviation value of each convolution layer feature extraction.

		Conv1	Conv2	Conv3	Conv4	Conv5
Short skirt	Mean	15.13	14.12	13.40	9.51	2.60
	Standard deviation	22.26	16.86	17.39	10.53	4.49
Trouser	Mean	19.90	6.44	8.90	4.81	0.98

Continued

	Standard deviation	81.37	20.95	19.47	11.82	5.59
Basketball shoes	Mean	18.70	6.15	9.58	6.29	1.55
	Standard deviation	57.17	20.44	21.77	14.54	7.09

From conv1 to conv5, the mean and standard deviation value of each category are gradually decreasing. This means that the feature information of images are extracted in a stable fashion. From conv3 to conv5, the standard deviation value of short skirt is less than trouser, and trouser is less than basketball shoes. The smaller the standard deviation value, the better the effect of feature extraction and the more stable the image feature.

5. Comparative Experiment Based on Shallow Convolutional Neural Network

In the application of modern technology, saving time cost and resource share rate are very important aspects that cannot be ignored. In a relatively simple task, such as collecting fewer images in the object, the shallow convolutional neural network can accomplish the task better, why should we design a complex network with higher time cost?

5.1. Image Preprocessing

There were 4000 images in our database and each product has 200 images in which 150 images were used for training purposes and 50 images for testing purposes. The image sizes are not the same. To facilitate the experiment, all images are normalized into $256 \times 256 = 65,536$ pixels.

Because of the small number of samples and the shallow network layers, this paper focuses on image preprocessing. In order to eliminate the influence of complex background on the network, a more intuitive method is to extract the recognition object from the image and then use the extracted region for training. It is necessary to detect the target object in the image, and the RCNN algorithm is the classical algorithm in deep learning for detecting target object. The RCNN algorithm was proposed by Girshick [23] in 2014 and achieved great success. The detection rate on PASCALVOC database was greatly increased from 35.1% to 53.7%.

Although RCNN has achieved good results, there are some obvious shortcomings, such as the number of bounding boxes is too large, the training time is long, and many bounding boxes overlap each other, resulting in repeated calculation. To solve these problems, an improved Fast-RCNN [24] has been proposed. The biggest difference between Fast-RCNN and RCNN is that the Fast-RCNN maps all bounding regions to the last convolution layer of the network, and then uses a ROI pooling layer to unify the sizes of different bounding regions. Only one feature extraction is needed for an image, and feature extrac-

tion is not performed for each bounding region, thereby greatly improving the efficiency of calculation.

Although the speed of Fast-RCNN is greatly improved compared to RCNN, there is still a need to optimize the large number of bounding regions. In view of this, the Faster-RCNN [25] algorithm is proposed. Faster-RCNN is characterized by extracting bounding regions from feature maps after the convolution layer rather than from the original image, so a Region Proposal Networks (RPN) is added to generate bounding regions based on Faster-RCNN.

This paper used Faster-RCNN to detect the location of clothing in the image, and then the image is normalized to 64×64 as input image.

5.2. Fine Tune and Training

Convolution neural network as a deep learning network structure requires a lot of data for training and a deep network structure in order to achieve better classification impacts. The training result based on the small samples and shallow convolutional neural network is often unsatisfactory. In view of this situation, this paper uses ImageNet database, which consists of 1.2 million images and 1000 categories for the shallow network pre-training. Network training is a process to update the initialization parameters to the optimal parameters. When the ImageNet training is completed, the trained parameters are stored in the shallow convolutional neural network. Then, input preprocessed product database for network training to obtain optimal parameters. The network storing the optimal parameters serves as a new shallow network model for feature extract. Finally, the Softmax is used to classify these features.

5.3. Shallow Convolutional Neural Network Model Architecture

Figure 7 shows the shallow convolutional neural network architecture and the trained parameters. The database was normalized to 64×64 RGB images after the preprocessing. The shallow network accepted 64×64 RGB images as input and output a vector for each block. It had one 8×8 convolution layer with a stride of 1, followed by a 3×3 max-pooling layer with a stride of 2. This was mapped into a 6×6 convolution layer, which increased the number of filters from 16 to 28. Next, a 3×3 max-pooling layer was mapped with a stride of 2, and the number of filters was 28. Following this, there were three 4×4 convolution layer with strides of 1. Finally, the feature maps were mapped into a 3×3 max-pooling layer with a stride of 2. Then, using softmax classifiers to classify 20 category product images. The network was trained using mini-batches with 25 images per batch and the training was set to 50 epochs, to speed up the gradient update with a learning rate set to 0.001.

5.4. Results and Discussion

Figure 8 shows the classification accuracy and cross entropy loss of the experiment on homemade database. To achieve the highest accuracy possible without
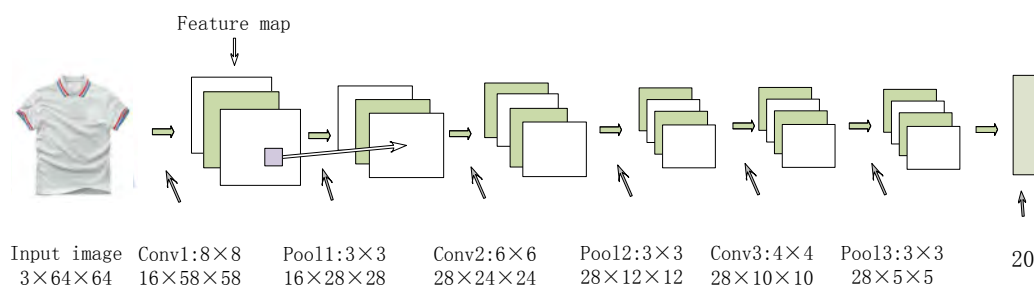
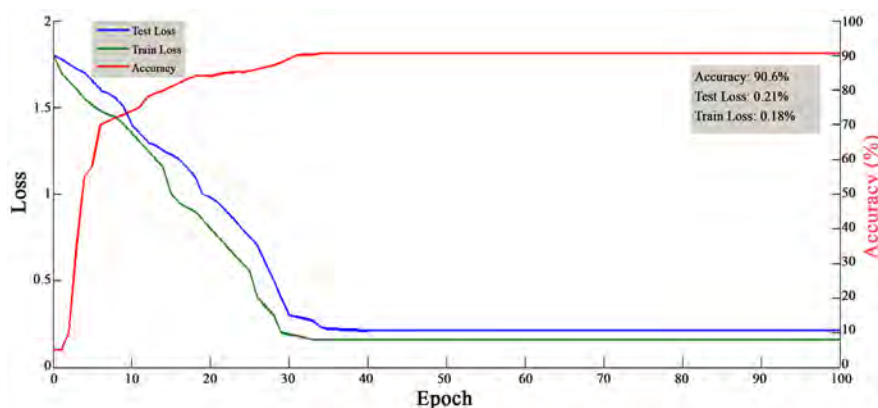**Figure 7.** Shallow convolutional neural network architecture.

Figure 8. Classification accuracy and cross entropy loss of experiment. Red line represents the test accuracy, blue line represents test loss, green line represents train loss.

overfitting the network, the training was set to 50 epochs. The average classification accuracy of the test was 90.6%. Setting appropriate learning rates in the experiment can improve the learning efficiency of network and therefore improve the classification accuracy. The learning rate was reduced three times before the experiment was stopped. At the beginning, we set the learning rate at 0.001. The test accuracy rapidly increased and the test loss rapidly declined. According to the decline of train loss curve, the learning rate of the network is relatively high. After 10 epochs, the test accuracy slowly increased, even decreasing, and the test loss was an upward trend. We therefore set the learning rate at 0.0005. It was observed that the test accuracy of the network increased again and the test loss slowly decreased. After 30 epochs, the test accuracy and test loss was not stable. We set the learning rate at 0.0001. It was observed that the test accuracy was high and the train loss continued to decline, then stabilized after 35 epochs.

Overall, the shallow convolutional neural network can save time cost and resource share rate by reducing network layers and training epochs. However, it is impossible to achieve high classification accuracy by simply reducing the number of network layers and iterations, which requires processing in image preprocessing and network initial parameter modulation.

6. Conclusions

In this study, we designed and trained a feature-based deep CNN for color image

classification in e-commerce domains, which are comprised of data augmentation pre-processing, feature extraction, and softmax classification. The proposed network is feasible and effective by evaluating the filter capacity and coverage of the network. To evaluate the classification performance of this technique, experiments were conducted using a homemade product image database taken from shopping websites on a total of 20,000 color images, with an average accuracy of 92.1%. Empirical results for the image database have shown that the proposed feature-based deep CNN is very competitive when compared with traditional content-based image classification for all performed experiments.

To solve the problem of long training time and high resource share of deep convolutional neural network, this paper designed a shallow convolutional neural network to achieve the classification accuracy of 90.6%.

The proposed network fine-tunes the parameters and architecture based on CNNs (as reported in this study) can be readily integrated and implemented in other image recognition and classification domains.

The potential future work involves improving new and deeper network architectures for product image classification; applying the CNN on other image databases and improving the classification accuracy by transfer learning.

Data Availability

The data used in the numerical analysis are mainly obtained from Internet-based e-commerce databases, including T-mall, Jingdong, and Amazon.

Funding Statement

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Zhou, X.S. and Huang, T.S. (2002) Unifying Keywords and Visual Contents in Image Retrieval. *Multimedia IEEE*, **9**, 23-33.

[2] He, R., Xiong, N., Yang, L.T., *et al.* (2011) Using Multi-Modal Semantic Association Rules to Fuse Keywords and Visual Features Automatically for Web Image Retrieval. *Information Fusion*, **12**, 223-230.

[3] Xu, J. and Shi, P.F. (2004) Active Learning with Labeled and Unlabeled Samples for Content-Based Image Retrieval. *Journal of Shanghai Jiaotong University*, **38**, 2068-2072.

[4] Jia, S.J., Kong, X.W., Fu, H., *et al.* (2010) Product Images Classification with Multiple Feature Combination. *Proceedings of the* 1*st International Conference on E-Business Intelligence* (*ICEBI* 2010), Atlantis Press, 446-469.

[5] Nilsback, M.E. (2009) An Automatic Visual Flora-Segmentation and Classification

of Flower Images. Oxford University, Oxford.

[6] Yao, B., Khosla, A., Li, F.F., *et al.* (2011) Combining Randomization and Discrimination for Fine-Grained Image Categorization. *Computer Vision and Pattern Recognition IEEE*, Colorado Springs, 20-25 June 2011, 1577-1584.

[7] Yao, B. and Khosla, A. (2012) Codebook-Free and Annotation-Free Approach for Fine-Grained Image Categorization. *Computer Vision and Pattern Recognition IEEE*, Providence, 16-21 June 2012, 3466-3473.

[8] Krause, J., Stark, M., Jia, D., *et al.* (2014) 3D Object Representations for Fine-Grained Categorization. *International Conference on Computer Vision Workshops IEEE*, Sydney, 2-8 December 2013, 554-561.

[9] Dyrmann, M., Karstoft, H., Midtiby, H.S., *et al.* (2016) Plant Species Classification Using Deep Convolutional Neural Network. *Biosystems Engineering*, **151**, 72-80.

[10] Sun, Y., Liu, Y., Wang, G., *et al.* (2017) Deep Learning for Plant Identification in Natural Environment. *Computational Intelligence and Neuroscience*, 2017, Article ID: 7361042.

[11] Krizhevsky, A., Sutskever, I., Hinton, G.E., *et al.* (2012) ImageNet Classification with Deep Convolutional Neural Networks. *International Conference on Neural Information Processing Systems*, Lake Tahoe, 3-6 December 2012, 1097-1105.

[12] Russakovsky, O., Deng, J., Su, H., *et al.* (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, **115**, 211-252.

[13] Lecun, Y., Bengio, Y., Hinton, G., *et al.* (2015) Deep Learning. *Nature*, **521**, 436.

[14] Rumelhart, D.E., Hinton, G.E., Williams, R.J., *et al.* (1986) Learning Representations by Back-Propagating Errors. *Nature*, **323**, 533-536.

[15] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition.

[16] He, K., Zhang, X., Ren, S., *et al.* (2016) Deep Residual Learning for Image Recognition. *Computer Vision and Pattern Recognition*, Las Vegas, 770-778.

[17] Ioffe, S. and Szegedy, C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Computer Science*, 448-456.

[18] Dieleman, S., De Fauw, J., Kavukcuoglu, K., *et al.* (2016) Exploiting Cyclic Symmetry in Convolutional Neural Networks. 1889-1898.

[19] Mount, J. (2011) The Equivalence of Logistic Regression and Maximum Entropy-models. http://www.win-vector.com/dfiles/LogisticRegressionMaxEnt.pdf

[20] Glorot, X., Bordes, A., Bengio, Y., *et al.* (2011) Deep Sparse Rectifier Neural Networks. *International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, 315-323.

[21] Cao, X. (2015) A Practical Theory for Designing Very Deep Convolutional Neural Networks Classifier Level. Technical Report.

[22] Luo, W., Li, Y., Urtasun, R., *et al.* (2016) Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, Barcelona, 5-10 December 2016, 4898-4906.

[23] Girshick, R., Donahue, J., Darrel, T., *et al.* (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587.

[24] Girshick, R. (2015) Fast-RCNN. *Proceedings of the IEEE Conference on Computer Vision*, Santiago, 7-13 December 2015, 1440-1448.

[25] Ren, S., He, K., Girshick, R., *et al.* (2015) Faster-RCNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems*, Montreal, 7-12 December 2015, 91-99.

Nomenclature

B: trainable scale for network layer

γ: trainable bias for network layer

μ: mean of image batch

σ: standard deviation of image batch

$\times$: image batch

ReLU: the rectified linear unit
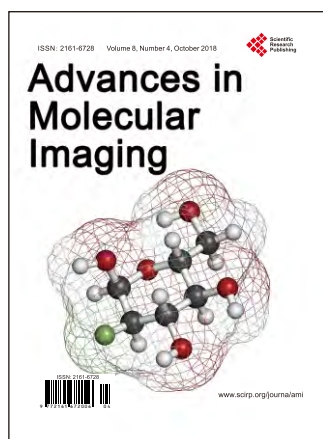
GIST: Global descriptor

PHOG: Pyramid Histogram of Orientated Gradients

PHOW: Pyramid Histogram of Words

LBP: Local Binary Pattern

HOG: Histogram of Gradient Orientations

SIFT: Scale Invariant Feature Transform

Call for Papers

Advances in Molecular Imaging (AMI)

ISSN 2161-6728 (Print)   ISSN 2161-6752 (Online)

http://www.scirp.org/journal/ami

Advances in Molecular Imaging is a peer-reviewed, open access journal that publishes original research articles, review articles, case reports, and clinical studies in all areas of molecular imaging:

Editor-in-Chief

Prof. Orhan Nalcioglu   University of California-Irvine, USA

Editorial Board

Subject Coverage

Chemical Synthesis Platform Technologies
Computer Vision and Image Understanding
Contrast Media & Molecular Imaging
Imaging & Microscopy
Imaging Decisions MRI
Imaging Systems and Technology
ImmunoPET and ImmunoSPECT
In Vitro Molecular Diagnostics (IVMD)
Molecular Diagnostics in Cancer and Immune Disorders

Molecular Imaging and Biology
Molecular Imaging Using ImmunoPET
MRI/MDCT/PET Imaging of Various Organs
Neuroimaging
Nuclear Medicine and Molecular Imaging
Onco-Imaging
Preclinical Imaging Systems
Systems Biology
Tracer and Pharmacokinetic Modeling

Notes for Intending Authors

We are also interested in: 1) Short reports—2-5 page papers in which an author can either present an idea with a theoretical background but has not yet completed the research needed for a complete paper or preliminary data; 2) Book reviews—Comments and critiques.

Website and E-Mail

http://www.scirp.org/journal/ami   E-mail: ami@scirp.org

What is SCIRP?

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

What is Open Access?

All original research papers published by SCIRP are made freely and permanently accessible online immediately upon publication. To be able to provide open access journals, SCIRP defrays operation costs from authors and subscription charges only for its printed version. Open access publishing allows an immediate, worldwide, barrier-free, open access to the full text of research papers, which is in the best interests of the scientific community.

- High visibility for maximum global exposure with open access publishing model
- Rigorous peer review of research papers
- Prompt faster publication with less cost
- Guaranteed targeted, multidisciplinary audience