HA NOI UNIVERSITY OF SCIENCE AND TEHNOLOGY
SCHOOL OF ELECTRICAL AND ELECTRIONICS

REPORT
TECHNICAL WRITING AND PRESENTATION

Topic:

FEATURE DETEDETECTION AND DESCRIPTION FOR ENDOSCOPY FEATURE EXTRACTION : A SERVEY

Student : ĐOÀN QUANG LƯU

Class : ET-E9 02 - K65

Lecturer : DR. NGUYỄN TIẾN HÒA

Hà Nội, April 5, 2023

ABSTRACT

Automated analysis of the gastric lesions in endoscopy videos is a challenging task and dynamics of the gastrointestinal environment make it even more difficult. In computer-aided diagnosis, gastric images are analyzed by visual descriptors. Various Deep Convolutional Neural Network (DCNN) models are available for representation learning and classification. Besides, using feature descriptor is also a great performance method. Some state-of-the-art method includes GLCM, BDIP, BVLC, etc. In this report, a various researches will be discussed.

Phân tích tự động các tổn thương dạ dày trong các video nội soi là một nhiệm vụ đầy thách thức và tính năng động của môi trường đường tiêu hóa càng khiến nhiệm vụ này trở nên khó khăn hơn. Trong chẩn đoán có sự hỗ trợ của máy tính, hình ảnh dạ dày được phân tích bằng các mô tả trực quan. Nhiều mô hình Mạng thần kinh chuyển đổi sâu (DCNN) khác nhau có sẵn để học biểu diễn và phân loại. Bên cạnh đó, sử dụng bộ mô tả tính năng cũng là một phương pháp hiệu suất tuyệt vời. Một số phương pháp hiện đại bao gồm GLCM, BDIP, BVLC, v.v. Trong báo cáo này, nhiều nghiên cứu khác nhau sẽ được thảo luận .

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATAIONS

GLCM	Gray Level Co-occurence Matrix
BDIP	Block Difference of Inverse Probabilities
BVLC	Block Variation of Local correlation Coefficients
CSI	Channel State Information
HBFE	Histogram-Based Feature Extraction
CH	Co-occurrence Histogram
CVM	Support Vector MMachine
K-NN	K Nearest Neighbor
GF	Gabor Filters
LBP	Local Binary Pattern
HLAC	Higher order Local Auto-correlation

LIST OF FIGURES

1 Introduction

The human digestive tract consists of a hollow muscular tube that begins in the oral cavity and continues through the pharynx, esophagus, stomach, and intestines to the rectum and anus. Generally, it is about nine meters long and is divided into upper and lower zones. Digestive disorders including constipation, irritable bowel syndrome, hemorrhoids, anal fissures, anal abscess, anal fistula, perianal infections, diverticulitis, colitis, colon polyps and cancer . Many of these can be minimized or prevented by maintaining a healthy lifestyle, practicing good bowel habits, and having cancer screening tests. Besides, this can greatly aid medical diagnosis and reduce the time and cost of investigative procedures. Therefore, evaluation endoscopy is the gold standard for identifying abnormalities of the human gastrointestinal tract.

Nonetheless, automated analysis of gastric lesions in endoscopic videos is challenging .Automated identification of gastrointestinal abnormalities with endoscopic images is a challenging task even for experienced gastroenterologists which could greatly aid medical diagnosis and reduce the time and cost of investigational procedures. in addition because the dynamics of the gastrointestinal environment make this task more difficult. In computer-aided diagnosis, gastric images are analyzed using visual descriptions. In computer-aided diagnosis, gastric images are analyzed using visual descriptions.

Deep convolutional neural network (DCNN) is currently an active research subject in health care and it has a potential to assist medical experts to deliver quality care at a large scale (Riegler et al., 2016; Topol, 2019) . Various DCNN models are available for representation learning and classification. Besides, using feature descriptors is also a great performance method.

Therefore, we have studied several modern automated gastric lesion analysis methods. Based on that, it is possible to analyze the method's characteristics and suitability for different specific cases. Some modern methods include GLCM, BDIP, , BVLC, etc. In this report, different studies will be discussed.

2 Feature Detection and Descriptor

2.1 Image Processing

2.1.1 Image Registration

Image registration is defined as a process that overlays two or more images from various imaging equipment or sensors taken at different times and angles, or from the same scene to geometrically align the images for analysis.

2.2 Vector Representation

2.2.1 Histogram-based feature extraction

Since, Histogram-based feature extraction (HBFE) have been used in various researches, in this survey, 5 popular ways of using histogram are stated.

- Luminance histogram: The RGB color channels are combined into a 8 bpp luminance channel Y in the YUV color space model. Subsequently a 1-dimensional, 256 bin histogram is constructed using the luminance values.
- Color channel histograms: a 1-dimensional, 256 bin histogram is constructed for each of the three color channels RGB independently.
- Color histogram: a 3-dimensional histogram is constructed where the three color channels represent the three dimensions, each dimension using 256 bins. For each pixel, the entry in the histogram is incremented according to its respective R, G, and B value.
- Co-occurrence histogram (CH): The co-occurrence histogram [1] counts the number of pairs of pixels exhibiting specific color or luminance values that occur at certain separation distances in image space. Therefore, the CH adds geometric information to the classical color histogram, which abstracts away all geometry. By adjusting the distances over which we check cooccurrences, we can adjust the sensitivity of the algorithm. We use horizontal and vertical adjacency as separation distance and construct the CH out of the luminance channel Y. This results in a 2-dimensional histogram where the luminance values of the two adjacent pixels represent the two dimensions, each with 256 bins.
- Using CH to compute further information of a picture (see Equation). 2

$$Uniformity = \sum_{i,j} p_{i,j}^2 \tag{2.1}$$

$$Entropy = \sum_{i,j} p_{i,j} \log P(i,j) \tag{2.2}$$

$$Maximumporbability = max(p_{i,j}) \tag{2.3}$$

$$Constrast = \sum_{i,j} |i-j|^k (p_{ij})^2 \tag{2.4}$$

$$Inverse\ difference\ moment = \sum_{i,j} \frac{(p_{ij})^2}{|i-j|^k} \tag{2.5}$$

$$Correlation = \sum_{i,j} \frac{(i-u)(j-u)p_{ij}}{\sigma^2} \tag{2.6}$$

The problem of using histogram vector is that histogram feature vector is not a vector in Euclidean Space. Therefore a histogram distance formula is proposed as the intersection between two distribution. Consider two distribution histogram $H(I)$ and $H(I\prime)$ calculated from two images I and $I\prime$. Since $H(I)$ and $(I\prime)$ have the same bin n, the normalized histogram is calculated in the Equation 2.7:

$$H(I) \cap H(I\prime) = \sum_{j=1}^{n} min(H_j(I), H_j(I\prime)) \tag{2.7}$$

2.2.2 Wavelet-based Feature Extraction

The (one-dimensional) DWT operates on a real-valued vector x of length 2^n, $n \in 2,3,...$, and results in a transformed vector w of equal length. Figure 3(a) and 3(b) illustrate the first two steps of the DWT for a vector of length 16. First, the vector x is filtered with some discrete-time, low-pass filter (LPF) h of given length at intervals of two, and the resulting values are stored in the first eight elements of w (see Figure 3). Otherwise, the low-pass filtered part of w (first eight elements for this example) can be further transformed using the identical procedure as outlined above. (see Figure 3) Filter coefficients for the low, high pass filter vector denoted h, g respectively, should be related in Equation 2.8

$$g_k = (-1)^k h_{n-k-1}, k \in 0,\ldots,n-1 \tag{2.8}$$

where n denotes the length of the filter (see Equation 2.9 and 2.10).

$$h = c_0 c_1 \Rightarrow g = c_1 - c_0 \tag{2.9}$$

$$h = c_0 c_1 c_2 c_3 \Rightarrow g = c_3 - c_2 c_1 - c_0 \tag{2.10}$$

The simplest wavelet filter is the Haar filter (see Equation 2.11)

$$h = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \tag{2.11}$$

Another very popular set of wavelet filters is due to Daubechies. The most compact of these has four coefficients (Daubechies-4), where h is given by:

$$h = \frac{1+\sqrt{3}}{4\sqrt{2}} \frac{3+\sqrt{3}}{4\sqrt{2}} \frac{3-\sqrt{3}}{4\sqrt{2}} \frac{1-\sqrt{3}}{4\sqrt{2}} \tag{2.12}$$

Computing the two-dimensional DWT is conducted through repeated application of the one-dimensional DWT. Figure **??** illustrates the basic, one-level, two-dimensional DWT procedure. First, a one-level, one-dimensional DWT is applied along the rows of the image. Second, another one-level, one-dimensional DWT is applied along the columns of the transformed image from the first step. As depicted in Figure 3 (left), the result of these two sets of operations is a transformed image with four distinct bands: (1) LL, (2) LH, (3) HL and (4) HH. Here, L stands for low-pass filtering, and H stands for high-pass filtering. The LL band corresponds roughly to a down-sampled (by a factor of two) version of the original image. The LH band tends to preserve localized horizontal features, while the HL band tends to preserve localized vertical features in the original image. Finally, the HH band tends to isolate localized high-frequency point features in the image.

2.2.3 Gray Level Co-occurence Matrix

Features extraction technique of GLCM is employed for textural information. It is a technique that allows for the extraction of statistical information from the image regarding the pixel pair distributions. GLCM is the most popular method for texture analysis and were demonstrated to feature a potential for effective texture discrimination. GLCM is defined by

$$P_d[i,j] = n_i j \tag{2.13}$$

where n_{ij} is the number of pair pixel occurrences (see Figure 3). It is computed by defining a direction, θ and a distance, d. Values (i,j) lying at distance d in the image. Then a pairs of pixels separated by d, computed across the defined θ are analyzed. As suggested in [2], directions of 0°, 45°, 90°, and 135° and the distance of 1 pixel is used in this research. The directions are chosen by assuming that the matrix is symmetrical. Symmetrical means the pixel pair is separated by d and –d. So, the count will consider the pixel in the direction opposite to d. Each pixel pair is computed to get a co-occurrence matrix. There are fourteen textural feature can be gathered from GLCM.

$$\textit{Angular Second Moment: } f_1 = \sum_i \sum_j \{p(i,j)\}^2 \tag{2.14}$$

$$\textit{Contrast: } f_2 = \sum_{n=0}^{N_g-1} \{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \} \, | \, p(i,j) \, (|i-j| = 1) \tag{2.15}$$

$$\textit{Correlation: } f_3 = \frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{2.16}$$

$$\textit{Sum of Squares - Variance: } f_4 = \sum_i \sum_j (i-\mu)^2 p(i,j) \tag{2.17}$$

$$\textit{Inverse Difference Moment: } f_5 = \sum_i \sum_j \frac{1}{1+(i+j)^2} p(i,j) \tag{2.18}$$

$$\textit{Sum Average: } f_6 = \sum_{i=2}^{2N_g} i p_{x+y}(i) \, where \, p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \, D[k] = \{a \in N^* | 2 \le a \le 2N_g\}, i+j=k \tag{2.19}$$

$$\textit{Sum Variance: } f_7 = \sum_{i=2}^{2N_g} (i-f_8)^2 p_{x+y}(i) \tag{2.20}$$

$$\textit{Sum Entropy: } f_8 = -\sum_{i=2}^{2N_g} p_{x+y}(i) log\{p_{x+y}(i)\} \tag{2.21}$$

$$\textit{Entropy: } f_9 = -\sum_i \sum_j p(i,j) log\{p(i,j)\} \tag{2.22}$$

$$\textit{Difference Variance: } f_{10} = Var[p_{x-y}], \, where \, p_{x-y}(k) = p_{x+y}(k) \, for D[k] = \{a \in N^* | 0 \le a \le N_g - 1\} \, |i-j| \tag{2.23}$$

$$\textit{Difference Entropy: } f_{11} = -\sum_{i=0}^{N_g-1} p_{x-y}(i) log\{p_{x-y}(i)\} \tag{2.24}$$

$$\textit{Difference Entropy: } f_{11} = -\sum_{i=0}^{N_g-1} p_{x-y}(i) log\{p_{x-y}(i)\} \tag{2.25}$$

$$\textit{Information Measure of Correlation (1): } f_{12} = \frac{H[XY] - H[XY1]}{max\{H[X], H[Y]\}} \tag{2.26}$$

$$\textit{Information Measure of Correlation (2): } f_{13} = (1 - \exp[-2.0(H[XY2] - H[XY])])^{1/2} \tag{2.27}$$

$$\textit{Maximal Correlation Coefficient: } f_{14} = (Secondlargesteigenvalue of Q)^{1/2} \tag{2.28}$$

where $p(i,j)$ is $(i,j)th$ entry in a normalized gray-tone spatial dependence matrix, $= P(i,j)/R$. $p_x(i)$ is ith entry in the marginal-probability matrix obtained by summing the row of $p(i,j)$, $= \sum_{j=1}^{N_g} P(i,j)$. N_g is the number of distinct gray levels in the quantized image. In Equation 2.26 and 2.27, $H[X,Y] = -\sum_i \sum_j p(i,j) log(p(i,j))$, $H[XY1] = -\sum_i \sum_j p(i,j) log(p_x(i) p_y(j))$, and $H[XY2] = -\sum_i \sum_j p_x(i) p_y(j) log(p_x(i) p_y(j))$. in Equation 2.28, $Q(i,j)) = \sum_k \frac{p(i,k) p(j,k)}{p_x(i) p_y(j)}$.

2.3 Feature Map Representation

2.3.1 Higher order Local Auto-correlation (HLAC)

The HLAC features [3] , F, represent the expressed characteristics for a complete image, derived from the product-sum operations of the following formula that represents auto-correlations(see Equation 2.29).

$$F_n = \sum_r I(r) \cdot I(r+a_i) \ldots I(r+a_n) \tag{2.29}$$

where N is the order of auto-correlation, r is the $X-Y$ coordinate vector of the colonoscopy image, $I(r)$ is the pixel value at r, and $a_i(i=1\ldots M)$ are displacement vectors around r. 25 masks are formed by the configurations of r and a_i, restricted by the second orders of autocorrelation $(N=0,1,2)$ within a 3x3 area around r.

2.3.2 *Integral image techniques*

In integral image techniques [4], The value G at any $X-Y$ coordinate $r(x,y)$ in the integral feature table is the sum of all the auto-correlations f at r (see Equation 2.30)

$$G(r(x,y)) = \sum_{x\prime\leq x}\sum_{y\prime\leq y} f(r(x\prime,y\prime)) \tag{2.30}$$

2.3.3 *Local binary pattern (LBP)*

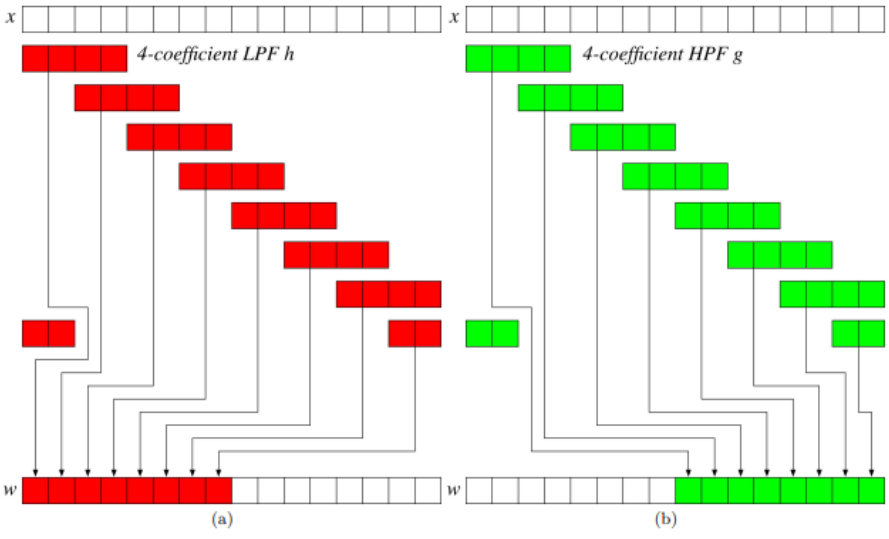
Local binary pattern (LBP) texture operator, which was first proposed by Ojala et al. The LBP feature is calculated as shown in Equation 2.31.

$$L(x,y) = I(x+p_i, y+p_i)\cdot D \tag{2.31}$$

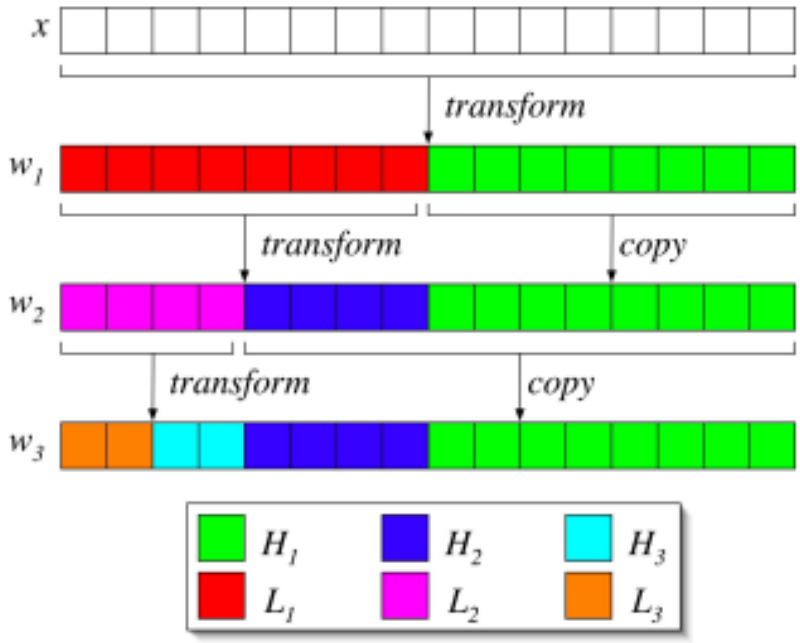
where $L(x,y)$ represent the LBP value at position (x,y), $I(x+p_i,y+p_i)$ is the intensity vector of size 8, representing the intensity of 8 cell around the center. D is the binomial weights vector $(2468\ldots^T)$

3 Figure

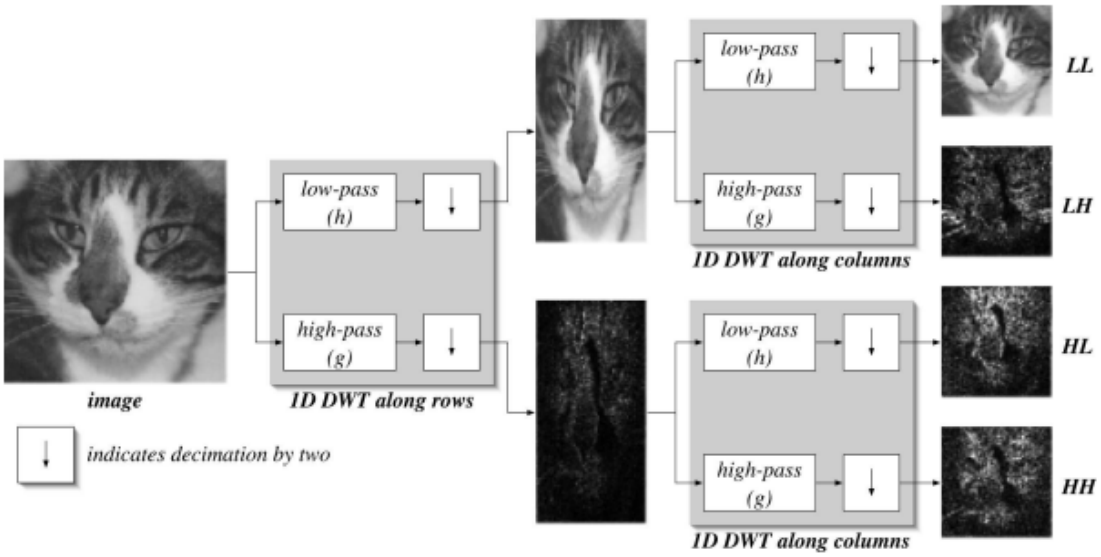
Wavelet Transform step. (a) First step of the DWT for a signal of length 16. (b) Second step of the DWT.



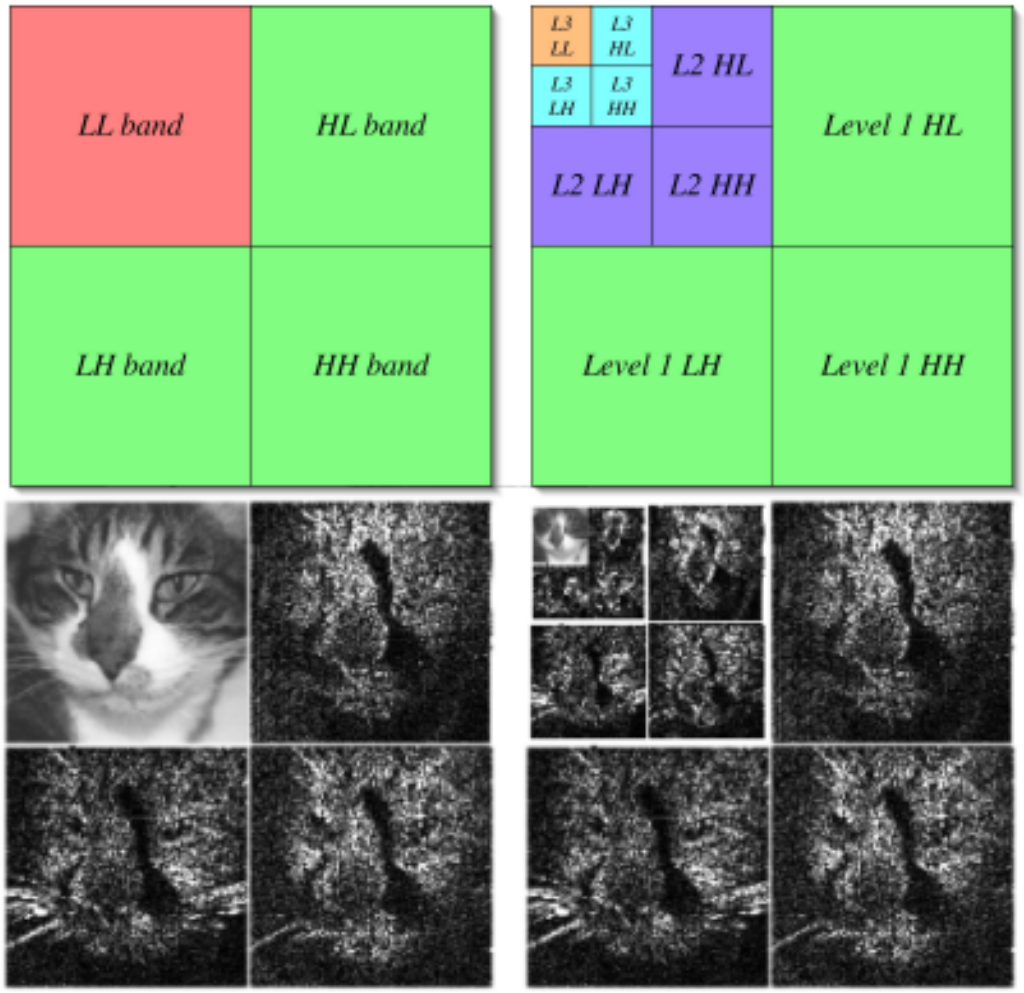
Three-level wavelet transform on signal x of length 16. Note that from w_1 to w_2, coefficients H_1 remain unchanged, while from w_2 to w_3, coefficients H_1 and H_2 remain unchanged.



One-level, two-dimensional DWT. First, the one-dimensional DWT is applied along the rows; second, the one-dimensional DWT is applied along the columns of the first-stage result, generating four sub-band regions in the transformed space: LL, LH, HL and HH .



Two-dimensional wavelet transform: (left) one-level 2D DWT of sample image, and (right) threelevel 2D DWT of the same image. Note that the LH bands tend to isolate horizontal features, while the HL band tend to isolate vertical features in the image.



(a) Image example, (b) construction of co-occurrence matrix (c) matrix framework for 0°, (d) matrix framework for 45°, (e) matrix framework for 90°, (f) matrix framework for 135°

0	0	1	2
0	2	2	3
2	3	3	2
1	1	2	3

(a)

i/j	0	1	2	3
0	#(0,0)	#(0,1)	#(0,2)	#(0,3)
1	#(1,0)	#(1,1)	#(1,2)	#(1,3)
2	#(2,0)	#(2,1)	#(2,2)	#(2,3)
3	#(3,0)	#(3,1)	#(3,2)	#(3,3)

(b)

2	1	1	0
1	2	2	0
1	2	2	4
0	0	4	2

(c)

2	0	0	0
0	0	1	2
0	1	6	1
0	2	1	2

(d)

2	0	2	0
0	0	2	1
2	2	0	6
0	1	6	0

(e)

0	0	2	1
0	0	1	1
2	1	2	2
1	1	2	2

(f)

4 Conclusion

Automated analysis of gastric lesions in endoscopic videos is essential. However, it is a challenging task even for experienced gastroenterologists. Deep convolutional neural networks (DCNNs) are currently an active research topic in the healthcare sector, and it has the potential to assist medical professionals in delivering quality care at scale. (Riegler et al., 2016; Topol, 2019). Various DCNN models are available for representation and classification learning. Also, using feature descriptors is also a great performance method.

Therefore, this report has investigated some modern automated gastric injury analysis methods. Based on that, it is possible to analyze the characteristics and suitability of the method for different specific cases. Some modern methods are studied GLCM, BDIP, , BVLC, etc. Methodologically it is essentially a process of overlaying two or more images from different devices or image sensors taken at different times and angles or from the same scene to align the images. then analyze the data from the images obtained with mathematical formulas , calculate with vectors ... and represent the feature histogram , in order to give the appropriate prediction results .

REFERENCES

[1] M. Y. M. H. R. Hussam Ali, Muhammad Sharif, "A shallow extraction of texture features for classification of abnormal video endoscopy frames," *Biomedical Signal Processing and Control*, 2022.

[2] F. R. F. B. S. M. D. R. M. T. Coimbra, "Invariant gabor texture descriptors for classification of gastroenterology images," *IEEE Xplore*, 2021.

[3] N. Otsu and T. Kurita, "A new scheme for practical flexible and intelligent vision systems," *" Proc. of IAPR Workshop on Computer Vision*, 1988.

[4] C. Franklin, "Summed-area tables for texture mapping," *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, 1984.