IIEEJ Transactions on Image Electronics and Visual Computing

Vol. 11, No. 1 2023

The Institute of Image Electronics Engineers of Japan

Editorial Committee of IIEEJ

Contact Information

The Institute of Image Electronics Engineers of Japan (IIEEJ)
3-35-4-101, Arakawa, Arakawa-ku, Tokyo 116-0002, Japan
Tel : +81-3-5615-2893 Fax : +81-3-5615-2894
E-mail : hensyu@iieej.org
http://www.iieej.org/ (in Japanese)
http://www.iieej.org/en/ (in English)
http://www.facebook.com/IIEEJ (in Japanese)
http://www.facebook.com/IIEEJ.E (in English)

IIEEJ Transactions on
Image Electronics and Visual Computing
Vol.11 No.1 June 2023

CONTENTS

Published two times a year by the Institute of Image Electronics Engineers of Japan（IIEEJ）
3-35-4-101, Arakawa, Arakawa-ku, Tokyo 116-0002, Japan
Tel：+81-5615-2893 Fax：+81-5615-2894 E-mail：hensyu@iieej.org http://www.iieej.org/

IIEEJ Invited Paper

Arithmetic Codes for Image Coding

: Their Advantages and the Effort to Make Them Practical

Fumitaka ONO (***Honorary Member***)

Tokyo Polytechnic University

<Summary> The arithmetic code has come to be used for image coding about 40 years ago. It was first adopted in bi-level image coding standard and nowadays widely adopted in multi-level image coding and video coding standards. Arithmetic code is classified as a non-block code in the information theory, and quite powerful by its ability and robustness for various target images. In this paper, its largest advantage for coding multi-context sources is described by comparing with the case using block codes. Additional advantage of arithmetic coding is its robustness based on the affinity with statistic learning. Therefore, the separation of model and entropy coding can be easily done and the current context and the observed symbol are enough for applying an arithmetic coding, while code set design and selection rule from the plural codes will be required in Huffman coding. Also, there were various efforts to make the arithmetic codes practical by referring to the designing parameters of known code including MELCODE.

Keywords: arithmetic code, Markov-model, entropy coding, international standard, MELCODE

1. Introduction

1.1 Background

The arithmetic code has come to be used for image coding about 40 years ago. It was first adopted in image coding standard about 30 years ago for JBIG. And the target has expanded from bi-level image standard to multi-level image and video standards.

Figure 1 shows the block diagram of image coding/decoding process. The image coding technology is composed of the image modeling and the entropy coding. The image modeling is the source conversion into the source having less entropy than that of the original. The quantization of the parameter of the converted source can be also included.

The entropy coding is the process of converting the observed event to the codeword having the corresponding information amount (entropy).

The entropy coding can be classified into block coding (table base) and non-block coding (procedure base) in information theory. The representative of the first type is Huffman coding and that of the second is the arithmetic coding.

Until recently, arithmetic coding was not introduced in the source coding chapter of the textbook of information theory for beginners. But as arithmetic coding has come to be adopted in various image coding standards in these days, such text book has come to mention about it as the understanding of its essence is quite important.

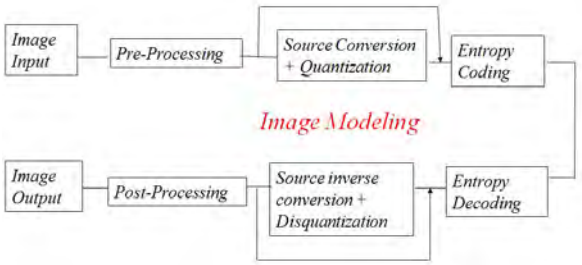


Fig.1 Block diagram of coder and decoder

In order to understand the reason why arithmetic code has been adopted so widely in international standards recently, I would like to describe about the advantage of the arithmetic coding by referring to the background of the initial adoption of the arithmetic code to the facsimile coding, and how researchers have made continuous effort to put arithmetic code more practical.

1.2 Basis of entropy coding

The entropy of the observed event having the probability $\boldsymbol{p}$ is calculated by $\log_2(1/\boldsymbol{p})$ bit. So the aim of entropy coding is how to assign codeword having $(\log_2(1/\boldsymbol{p}))$ bit for the event having probability $\boldsymbol{p}$.

The entropy coding process can be composed of

following three steps:

1) Recognition of the probability : Recognition of the context in connection with source modeling.
2) Estimation of the probability: Static estimation and dynamic estimation can be considered. The former is based on statistic data, and the latter will use such as Bayes estimation, state transition rule, etc.
3) Realization of codeword having ($\log_2(1/p)$) bit: Technique to assign appropriate codeword. The performance of 2) and 3) will correspond to the coding efficiency.

As stated above, the entropy coding technology can be classified into block coding (such as Huffman code) and non-block coding (such as arithmetic code). Huffman code is further classified to, a) Custom Huffman code, b) General Huffman code set, and c) Fixed Huffman code. The first one is the primary meaning of Huffman code but constructing Huffman code for individual source will not be so realistic. The second one can be regarded as true Huffman code if correct code selection algorithm is applied. The third one is most simple one but not robust depending on the target source.

2. Basic Knowledge of Arithmetic Coding

2.1 The origin of arithmetic coding

The origin of the arithmetic coding is considered as “Numerical line coding (NLC)]” to represent the area in the interval of [0,1] line, when the line is divided into areas of symbol sequences based on the probabilities. For example, consider a memoryless source with two symbols A and B, occurring with probability 0.5 and 0.5 respectively, as shown in **Fig.2**. If we assign “A” area to the bottom and “B” area to the top according to the ratio of the probability, arbitrary sequence starting with “A”, from such a source, will be assigned in the interval [0, 0.1](binary). Also the sequences starting with “AB” are assigned to the interval [0.01, 0.1] (binary) etc.

In NLC, to interpret the binary expressed address of the point in the final area, just below the decimal point, as the codeword sequence will be possible.

This idea was proposed by Peter Elias[1)], and his idea was introduced in the textbook “Information theory and coding” by Norman Abramson[1)]. Its efficient coding scheme proposed by P. Billingsley is also referred in the book.

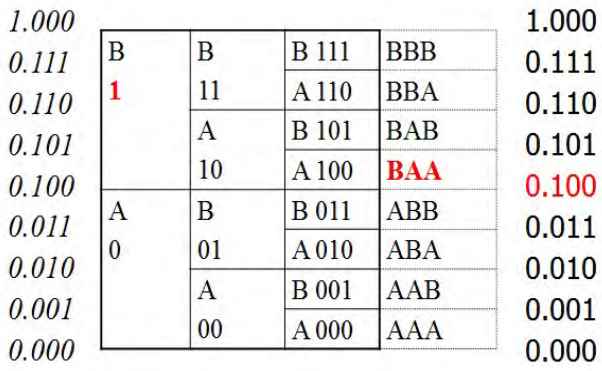


Fig.2 Most simple NLC example

If we use the code table below, the range of the address X (binary) will be,

if the first symbol is A, $0.0 \leqq X < 0.1$, and

if the first symbol is B, $0.1 \leqq X < 1.0$

The decoding of the first symbol can be done by comparing the binary number (corresponding to codeword sequence) with 0.1 (boundary address). So, the area of symbol sequence will be divided to upper half and bottom half of the numerical line, according to the fact whether first symbol is A or B. The area between 0.0and 0.1(binary) is for A, and the area between 0.1(binary) and 1.0 is for B.

2.2 Principle of arithmetic coding

The codeword in NLC can be expressed by its address, but the necessary number of address bits which corresponds to the length of codeword must be recognized. For example, the address of 0.1 and the address of 0.100 should be distinguished, as the codeword of the former is “1”, and the codeword of the latter is “100”. The address of 0.1 will mean “B”, and the corresponding area size is (1/2) because the probability of observing B as the first symbol is 0.5 and any sequence starting with B is allocated between 0.1 and 1.0, and the address of 0.1 will include all sequences starting with B. On the contrary, the address 0.100 will mean “BAA”, and the area size expressed by the address of 0.100 is 1/8 and will mean the probability of all sequences starting with BAA. Therefore it is important to consider the area size expressed by the address, which is the codeword itself. It should be noticed

that the area size of (1/8) will not guarantee that corresponding address can be expressed by 3bit. It will be discussed later in **2.4** but the difference of necessary bit length in the worst case is at most one.

The principle of the arithmetic coding is simply given by the following fact. The message of which codeword is n bit can be considered to be given the area of $(1/2)^n$ in the numerical line. The information theory tells that, the ideal codeword length n for the message having probability p is given by

$n = \log(1/p)$ bit

If the area size s is corresponding to n bit code, the minimum s will be

$s = (1/2)^n$

Then the ideal area size s for the message, of which probability is p, is given by

$s = 1/2^{\log(1/p)} = p \quad \therefore \quad s = p$

It is very simple result that ideal area size s in arithmetic coding is its probability p. Therefore, in NLC, the area of each symbol should be proportional to its probability[2].

2.3 General definition of arithmetic coding

In arithmetic coding, the symbol area will be divided according to its symbol probability in the numerical line [0,1), as encoding procedure will proceed.

If the final symbol in the sequence is given, the final area corresponding to the symbol sequence, in the numerical line [0,1), will be checked. A binary number γ, of which under decimal point bit is N, will express the area $[\gamma, \gamma+2^{-N})$. The minimum number of bits to be able to express the part included in the final area will be used as the codeword of the arithmetic coding.

2.4 Necessary code length in arithmetic coding[2]

Let the probability of symbol sequence as P, the necessary code length L_A in the arithmetic coding will be as following, because the assignment of symbol areas should follow alphabetical order (or predefined order) and rearranging is not allowed.

$$\log(1/P) \leqq L_A < \log(1/P) + 2$$

Figure 3 shows the example of the difference of the necessary codelength. If the final area size of B is 1/4 and it locates between 0.01 and 0.10, as left, the codeword of 01 ($0.01 \leqq x < 0.10$) can be used. But if it locates between 0.011 and 0.101, as right, three bit codeword (either 011 or 100) will be needed. This fact corresponds to the case mentioned in **2.2**.

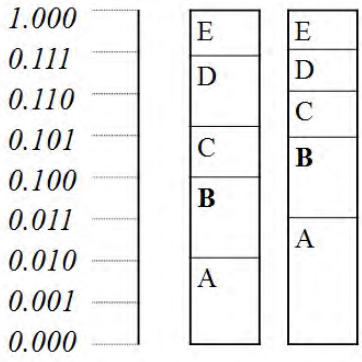


Fig.3 The difference of the necessary code length

In Huffman code, necessary code length L_B will be

$$\log(1/P) \leqq L_B < \log(1/P) + 1$$

By the way, in the block coding method restricting the binary order of codeword to follow symbols' alphabetical order (known as Gilbert-Moore coding[3]) necessary code length L_G will be

$$\log(1/P) \leqq L_G < \log(1/P) + 2$$

2.5 Renormalization and flush

a) Renormalization

As coding procedure progresses, the valid area, which is called as "augend" decreases. If the augend decreases less than half of the full augend size, the augend is doubled and the accuracy of the area of the symbol will be kept. This procedure is called "renormalization". If the augend is still less than half of the full size, such procedure will be repeated.

In binary arithmetic coding, MPS (More Probable Symbol) /LPS (Less Probable Symbol) recognition is convenient because the probability of MPS can be restricted to $\geqq 0.5$, and the probability of LPS can be restricted to $\leqq 0.5$.

If LPS occurs, renormalization will always happen, and if MPS occurs, renormalization will not always happen.

Figure 4 shows an example of the renormalization. Let the probability of MPS to be 3/4. If MPS will happen, the valid area will be decreased to 3/4, as shown in the area of right side described as the combination of (00,01,10), and the area will be further divided according to the probabilities of coming symbols. If LPS will happen, the valid area will become 1/4 of the original as painted yellow. Then the valid area shall be enlarged 4 times (2 more

bit in the address) to keep the necessary augend accuracy. In the Figure, as the top two bits of the yellow area is determined as “11”, we can put out the codeword of “11” to the channel and we can use the full area just as shown in left, for coming symbols.

b) Flush

Flush is the activity to put out the final code sequence finishing or quitting the coding procedure. In the type of flush, so-called minimum flush will put out the minimum code sequence to the channel for decoding the last symbol. There is a possibility that additional MPS will be decoded if the number of symbols is unknown to the decoder. By the way, such phenomenon is not due to the arithmetic coding procedure and it will also happen in block code with source extension too. In order to inform the number of symbols, using the control code beforehand is popular and other methods to devise the flush format is also possible.

2.6 Decoding procedure[4)]

At first, the lowest address and the highest address of the valid area will be set at the decoder. By comparing the boundary address of symbols and the address given by the received code sequence, initial symbol will be decoded. In the decoding register, the current area will be detected and if renormalization is needed we will double the area as required, and read the necessary bits of the code sequence into the decoding register. The decodability is guaranteed whatever code sequence will follow in the future. If renormalization is needed and there are no more data in the code sequence, we can know that we can stop decoding. At that time, it will happen that dummy MPSs can be decoded at the decoder. But it is not unique to arithmetic coding, since similar case will happen in block code which adopts source extension, too.

In some coding standards, the length of the code sequence may be restricted to the multiplication of eight. In that case, additional zero data will be added or primarily necessary zero data will be suppressed, if possible. In such a standard, the number of picture elements to be decoded must have been sent beforehand to let the decoder decode the correct number of events.

3. Markov-model Coding

From coding technology aspect, image is classified into three categories. They are 1) Bi-level images including text/line-art and halftone 2) Still colour images including natural photos, CG and compound images, and 3) Motion pictures including natural video, and animation (CG). The standardization and the application are also classified into these fields.

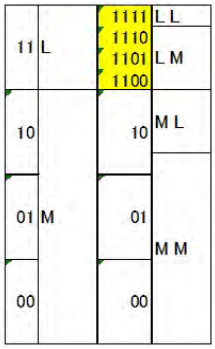


Fig.4 An example of renormalization

3.1 Bi-level image coding

For two-dimensional bi-level image coding, there were two ideas. The first one was two-dimensional Run Length coding[5)-8)], which utilizes the relation of two-dimensional edges. The other one is Markov-model coding[9)-13)] in which the conditional entropy of neighboring reference pixels will become the target entropy. The character and line-art images will be efficiently coded in both models, and in the second model, halftone images will be also efficiently coded.

In two-dimensional Markov-model coding, source extension of multi-context source is needed, which has not been theoretically treated, when this issue was studied world-wide about middle of 1970s, and various practical methods were proposed. We have made clear that the source extension of multi-context source can be done by either sequential extension or context-based extension[4)]. In sequential extension, the coming context has to be recognized beforehand based on the watched symbol, and the Huffman codes corresponding the combination of various MPS probability should be prepared. In context-based extension, codewords for different contexts will be created in parallel, and to keep the right order of the transmission, waiting for the completion of the codeword of other contexts and codeword interleaving will be needed. Also memory control of codeword data will be required under limited memory circumstances.

Table 1 Block type MELCODE

message	number of symbols	codeword
000・・・	M	1
1	1	m
01	2	
・	・	・
・	・	・
・	・	
000・・	M_1	m
000・・	M_1+1	m+1
・	・	・
・	・	・
000・・・	M	m+1

Table 2 Examples of Block type MELCODE

M=1		M=2		M=3		M=4	
message	codeword	message	codeword	message	codeword	message	codeword
0	1	00	1	000	1	0000	1
1	0	01	01	001	011	0001	011
		1	00	01	010	001	010
				1	00	01	001
						1	000

So, it can be said that context-based extension is more practical, but sequential extension is also possible if the Huffman code set to use is restricted by tolerating the degradation of the efficiency.

Our Markov-model coding method[10] using MELCODE[14] shown in Chapter 4 has achieved almost same performance compared with the best of two-dimensional RL coding methods[6),7)] in facsimile coding domestic standardization discussion.

Later, the nature of the arithmetic coding has been made clear[15)], and if we use arithmetic coding, coding procedure equivalent to sequential extension can be executed without worrying about complex code design needed in sequential extension when using block code, or codeword interleaving procedure needed in context-based extension when using block codes.

So, it can be said that multi-context source coding with source extension was elegantly solved by the use of arithmetic coding[2].

3.2 Multi-level image coding

Even in multi-level images, there will be a context which needs source extension to keep the coding efficiency. In order to utilize the study result of bi-level image coding, described in **3.1**, we have proposed the method of breaking down the multi-level sources into binary sources and using binary Huffman code to the binary sources[15)]. Though various breaking down method can be considered, N-level source will be divided into (N-1) binary sources to keep the entropy.

4. MELCODE

4.1 Block type MELCODE[14)]

MELCODE is a general binary Huffman code set proposed by us more than 40 years ago. In studying bi-level Markov-model image coding using context-based extension, general Huffman code set to be chosen for the probability model of the context is needed to develop. Let us call an integer parameter M as "order". In MELCODE, consecutive M MPSs will be coded by one bit codeword. Less than M MPS plus one LPS will be coded by fixed length code of ($m+1$) bit, if M is 2^m.

If M is not a power of 2, we will define m and M_1, satisfying the following formulas,

$2^{(m-1)} < M \leqq 2^m$

$M_1 = 2^m - M$

The code for any M is defined in **Table 1**. Some code examples for M can be found in **Table 2**.

In our study of binary Markov-model coding, context-based extension method will need a code set for group of the context, and MELCODE works very well for that purpose, as shown in **3.1**.

If switching the set of predefined orders for the distribution range of assumed probability, the minimum efficiency will be given at the switching point. In case of block-type MELCODE, if the order is limited to 2's power, the minimum efficiency will be given at the switching point of orders 1 and 2, and minimum efficiency will be about 0.959, as shown in **Fig.5**.

4.2 Arithmetic type MELCODE[16)]

After the discovery of the nature of arithmetic code, which is the convenience for Markov-model coding, we have tried to extend block type MELCODE to arithmetic type MELCODE, keeping the same efficiency in multi-context coding. It has been proposed by us more than 30 years ago, and has been adopted in several standards.

Block type MELCODE, when order $M=2^m$, can be considered as arithmetic code of constant LPS area regardless to the size of the augend. In **Fig.6**, the case of M=4, where LPS area width is 1/8

(=1/(2*M)), is shown. In single context coding, the augend will take one of (1, 7/8, 6/8, 5/8), and the LPS probability can be considered to alter as 1/8, 1/7, 1/6, and 1/5 according to the augend.

a) How to define arithmetic MELCODE

In block type MELCODE, the augend will take limited value, because of the context-based extension, and LPS area size will be always 1/(2*M) for M, when it satisfies the power of 2. In considering arithmetic MELCODE, as sequential source extension of multi-context sources will be accompanied, the augend will take any value between 1/2 and 1. Therefore the LPS area size corresponding to any augend size should be defined.

For the case of M=4, the augend is restricted to one of (1, 7/8, 6/8, 5/8) in single context coding and LPS area size is always 1/8. So, it will be natural to define LPS area size to be 1/8 if the augend is between 1 and 5/8. As the augend will not be less than 1/2, we have to define the LPS area size when the augend area is between 1/2 and 5/8. If we consider the case of the augend is 1/2, the LPS area size should be 1/16, because after the renormalization to follow, the augend size will be 1, and LPS area size will be 1/8. From this fact it will be natural to define LPS area size q as following if the augend size z will satisfy the following inequality.

$q=z/2-3/16$ if $1/2 < z < 5/8$

Generally when the order is M and the augend is z, the LPS area size q will be given by following rule.

$$q=(1/M_2) \quad \text{for} \quad z>\{1/2+(1/M_2)\}$$

$$q=(1/M_2)-(1/2)\{1/2-(z-1/M_2)\} \quad \text{for} \quad \{1/2+(1/M_2)\} \geqq z \geqq 1/2$$

where $M_2=2*M$

We are calling the operation when $\{1/2+(1/M_2)\} \geqq z$ as Over Half Processing (OHP) because the LPS area above the address of 0.1 (binary) should be made half. It should be noted that OHP will always accompany renormalization whether observed symbol is MPS or LPS. Therefore, when OHP occurs, the occurred symbol area (despite of MPS or LPS) must be at least doubled. So, the halving operation is actually not needed.

b) Introduction of OHP

OHP can be considered as the solution to the definition of arithmetic MELCODE for applying multi-context coding and also if non-integer order arithmetic MELCODE is introduced.

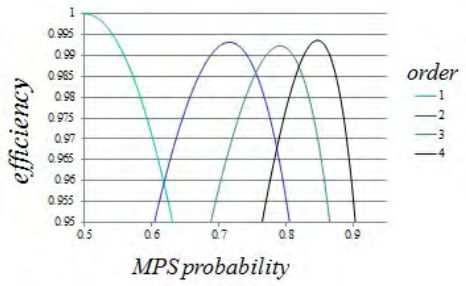


Fig.5 Coding efficiency of MELCODE

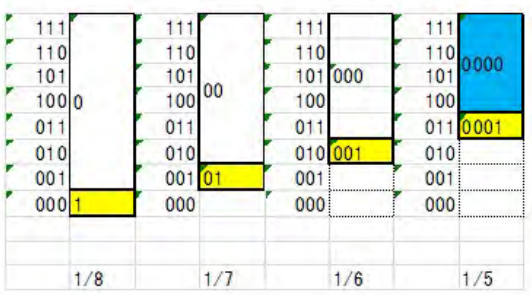


Fig.6 MELCODE of M=4 treated as arithmetic code

Concerning non-integer order arithmetic MELCODE, for example, order value of 1.5 can be considered between the order of 1 and 2. **Figure 7** shows the efficiency when the orders of 1, 1.5, 2 and 3 are used as a set, the minimum efficiency will be raised to about 0.985 which will be given at the switching point of the orders of 1 and 1.5.

By the way, for block-type MELCODE, of which order is not the power of 2, the LPS size is not constant. **Figure 8** shows the case when the order is 3, where symbol "0" means MPS, and symbol "1" means LPS. If we introduce arithmetic MELCODE of which order is 3 according to the definition given above, the LPS size will be 1/6, that is $1/M_2$, and the augend value in single context coding is shown as left. From Fig.8, it can be said that the efficiency will be same for block type coding (right) and for arithmetic coding (left) since LPS probability is one of (1/6,1/5,1/4) set in both cases.

4.3 Set of probability models

In block type MELCODE, the average efficiency or minimum efficiency can calculated for the set of the orders restricted to two's power. For the arithmetic MELCODE, to keep the consistency with block-type MELCODE, the order was limited to

two's power at first. However, since arithmetic MELCODE can define non-integer value of the order, it will be natural to design the set of orders from the viewpoint of minimum efficiency. Such optimization will be effective especially for large LPS probability case[17].

5. Design issues of Practical Arithmetic Codes

5.1 Design parameters of AC

Major design parameters of AC (Arithmetic Code) will be given as following.

1) Binary or M-ary: Binary is currently quite popular, and even for multi-level sources, usage of binary AC with the process of breaking down the M-ary source into binary sources is dominant.
2) Register size (bits): This is corresponding to symbol probability accuracy. Long register size is required for high skew data, such as high resolution facsimile. The register size of QM-coder[18] used in JBIG standard is 16 bit, and this will be considered to be satisfying.
3) Symbol area size assignment rule: This will be most important since this issue can be the main characteristic of the individual AC.
4) Carry control method: Various methods are proposed as described later.
5) Adoption of MPS/LPS in binary AC: MPS/LPS concept is convenient since the symbol probability range can be halved as probability of MPS will be $\geqq 0.5$ and the probability of LPS will be $\leqq 0.5$, with judging process of MPS/LPS. Which symbol of MPS/LPS should be allocated to the larger/smaller side of the numerical line is up to the designer of the system.
6) Number of probability models: For the given range of LPS probability, we will assume a set of model sources having representative values of an LPS probability. The target source will be approximated by one of the model sources. This parameter is related to the statistic coding efficiency, together with 3).
7) Renormalization timing: Renormalization will be done soon after the augend has decreased to less than half of the full augend size, but some AC will do the renormalization less often for simplification and also for speed-up

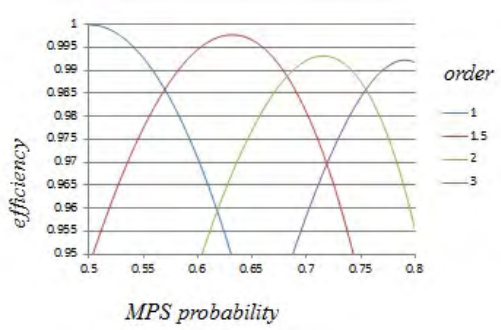


Fig.7 Introduction of no-integer order MELCODE

Fig.8 MELCODE of M=3:Block-type(left) and arithmetic type (right)

8) Probability estimation and update timing: To use state transition diagram is popular. This is related to the dynamic coding efficiency, together with 3) and 6).

5.2 Symbol area size assignment rule

In the AC, symbol area width will be ideally calculated by [valid area width (Augend) × probability]. Valid area width will be varying among 1/2 and 1, when maximum augend is defined as 1, under the usual usage of renormalization.

The multiplication procedure is not practical as the area size to be assigned should be digitalized into given accuracy of register size and it must be coincident between encoder and decoder. From that sense, the area size should be tabularized, and using the less table size will be more convenient.

The first practical arithmetic code is LR code[19], proposed by G. G. Langdon & J. J. Rissanen, from IBM. In LR code, LPS probability is limited to the power of (1/2). So, the area size of LPS can be calculated by shifting operation of the augend. The second practical arithmetic code is Q-Coder[20] also proposed by IBM. As the LPS size is fixed regardless of the size of the augend, this is called as the subtraction type arithmetic coding. MELCODE as described in **4.** is the combination of fixed size LPS

and OHP. QM-coder[18),21)] and MQ-coder[22)] are both adopting fixed size LPS and MPS/LPS conditional exchange. It can be said the performance of fixed size LPS will be degraded around the LPS's probability is 0.5, and MELCODE and QM/MQ has the effect of improving such case.

5.3 Carry over problem

For example, if the valid area covers both upper and lower side of 0.1, there is the possibility of the final code sequence will be 0.100000… or 0.0111111…, and the highest bit will not be determined for long time, and the delay to put out the codeword into the transmission line will happen. In order to solve this issue, various controlling methods have been studied as following. In all methods, the encoder and the decoder will take same procedure.

1) Bit stuffing method[20)]: Tentative result will be sent out not waiting the final result. The final result will be sent later using controlling code.
2) Moving the boundary method[23)]: To move the MPS/LPS boundary to just 0.1 compulsorily, sacrificing the loss of some coding efficiency.
3) Discard the smaller area method[23)]: To change the top address or bottom address intentionally to be 0.1 by discarding either of upper or lower area of 0.1, sacrificing the loss of some coding efficiency.

By the way, in order to create a set of control codes starting FF, zero-insertion will be taken place. This issue and carry over problem are primarily independent but bit stuffing will be done as a part of the former work.

5.4 Coding efficiency of binary AC

The multi-context source will be composed of multiple sources depending on the context.

For the distribution of LPS probability of multiple sources, we will prepare a set of model sources having representative values of an LPS probability. The target source will be approximated by one of the model sources. Therefore, the minimum efficiency for the assumed range of the LPS probability will be given at the switching point of two neighboring model sources.

For the distribution of LPS probability, we will assume a set of model sources having representative values of an LPS (MPS) probability. The target source will be approximated by one of the model sources. Therefore, the minimum efficiency for the given range of the LPS probability will be given at the switching point of two neighboring model sources. The multi-context source will be composed of multiple sources depending on the context.

In order to design the probability models, minimum efficiency can be first assumed. If the number of models is assumed first, to let the efficiency of the switching point same will be usual.

The static coding efficiency will be given by the curve of switching best of the model sources. Concerning the dynamic efficiency, state transition rule to move one of the model sources will be designed[20)-22)]. The number of states can be larger than the number of model sources. The timing of the transition of the states will be the appearance of LPS or MPS, and at least counter will be needed for MPS[24)]. The counter will be also necessary for LPS having larger probability. To utilize renormalization timing will be quite effective as in Q-coder[20)].

5.5 Basic comparison of AC and HC

As described in Chapter 4, basically AC can be considered as the extended form of HC (Huffman Code). In HC, the all symbol sequence is divided by the collection of messages, and the probability of each message is limited to be $(1/2)^n$, where n is an integer. On the contrary, in AC, there is no concept of message, and the probability of only total sequence will be limited to $(1/2)^n$, where n is an integer. Therefore, AC will approximate the probability of all symbol sequence by the power of (1/2), the efficiency will be theoretically higher than that of HC.

For the adaptability, the update timing of HC is limited to a break between messages, while AC can choose any timing, for example, timing when the symbol counter reaches predefined value, timing of renormalization, etc.. Concerning the separation of model and entropy coding, the current context and the observed symbol are enough to be given for AC. While code design and selection rule from the plural codes will be required in HC.

In HC, source extension is done through the definition of the message. As described in **3.1**, the sequential source extension of multi-context sources will ask to prepare enormous kinds of messages

because the context to follow will take various values of probabilities. So, the restriction of reference events (pixels in images) will be accompanied in many cases, and still it is difficult to design the appropriate codes. As for the context-based extension, extended events of the same context will have same probability model and set of codes corresponding to the number of probability models will be needed.

On the other hand, sequential source extension is possible by nature in arithmetic code, and the extension of total sequence can be also possible. By this nature, arithmetic coding can be the powerful solution for Multi-context source coding.

5.6 The well-known AC

The first arithmetic code applied to the Markov-model images (facsimile images) was LR code[19], and it has succeeded to let researchers know the potency of the arithmetic code for Markov-model sources. In the method, LPS probability is limited to the power of (1/2). So, the area size of LPS can be calculated by shifting operation of the augend.

The arithmetic coding standardization of JBIG and JPEG standards was took place in ISO and CCITT. For the purpose, Q-coder[20] (from IBM), arithmetic MELCODE[16] (from Mitsubishi) and MM-coder[25] (from AT&T) were proposed and QM-coder[18] was newly defined as the sole standard. In QM-coder, fixed size LPS area (by IBM) is modified by MPS/LPS conditional exchange[18] (by Mitsubishi) and 113 states transition rule [21] (by AT&T) was adopted. Later in standardizing JPEG2000 and JBIG2, MQ-coder[22] has been defined. In MQ-coder, 47 states transition rule was newly adopted. In JPEG LS standard part-2, arithmetic coding which is combining CJ-coder and MELCODE was adopted. For H.264, CABAC[26] was introduced, and the LPS area size in CABAC is fixed according to the 4 classes of the augend size.

6. Recent studies

6.1 STT-Coder (State Transition Table coder)

We have proposed STT-Coder[27] to improve the simplicity of the binary arithmetic codes. Its basic idea is the trial to make retaining the bottom address unnecessary, when renormalization is done. In other words, flush will be done whenever renormalization

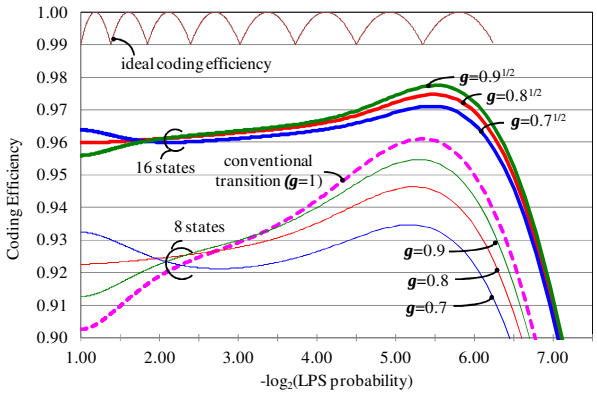


Fig.9 The performance of STT-coder

is needed. Therefore, the trial to avoid the loss in flush will be required. Such idea will make the arithmetic coder possible to be implemented by state transition table.

In studying STT-coder, a specific parameter called "offset" is introduced. This parameter shows the difference between the bottom address of the current valid interval and the address given by the already determined code bit stream. It was found that there is an optimum number of offset, which is four, and the best combination is (0, 16, 24, 28) in 6bit register system. The coding efficiency of 6bit system using 8 probability states shows 98.5% covering the range of MPS probability between 0.5 and 0.95.

For dynamic coding[28], a concept of "transit LPS ratio" was newly introduced especially for high LPS probability range, and using 16 states and 8 probability models, the efficiency of 96% was achieved as shown in **Fig.9**.

6.2 Multi-alphabet arithmetic coder[29]

The study of multi-alphabet arithmetic coder has not been so popular. For the arithmetic coding of multi-level sources, most popular method is to break down the multi-symbol source into binary symbol sources. The advantage of this method is to utilize the knowledge of binary arithmetic coding. The disadvantage is taking longer time than multi-level arithmetic coding.

In designing multi-symbol arithmetic coding, the first issue to be considered is source modeling. The most simple modeling will be to assume the probability ratio of Nth and (N+1)th symbol $\boldsymbol{R}$ to be constant. Then we assume single parameter for modeling. Concerning the calculation of area size of

symbols, areas of 2nd to last symbols to be fixed independent to the augend will be most practical. So we call such arithmetic code as "Fixed type arithmetic code".

If its efficiency is not satisfactory, classifying the range of augend into two or more categories will be necessary (for example, two tables will be prepared for the smaller augend, which is between 0.5 and say 0.7, and larger augend, which is between say 0.7 and 1.0).

We assume that the probability ratio of Nth and (N+1) th symbol to be constant as R. If the number of symbols is enough large, the probability of the top symbol will be (1-R).

Table 3 shows the comparison of coding efficiency with Huffman code not supposing source extension when the number of symbols is 256. In the Table, for example, if R=0.245, the probability of top symbol will be 0.755 and the advantage of the arithmetic coding over Huffman will be remarkable because of source extension effect in arithmetic coding. For larger R, the efficiency of arithmetic coding will be not so good as Huffman because of the effect of fixed area allocation of arithmetic coding. The "double arithmetic" means the augend is classified to two categories and the "single arithmetic" means that the augend is not classified. As the improvement of double arithmetic from single arithmetic is not so large, single arithmetic sounds enough.

Table 3 Coding efficiency of multi-level arithmetic code

	Huffman	Arithmetic	Arithmetic
R	14bits	Single table	Double table
0.841	0.988	0.946	0.971
0.707	0.985	0.958	0.973
0.5	0.986	0.964	0.973
0.382	0.955	0.964	0.97
0.293	0.871	0.961	0.966
0.245	0.803	0.959	0.963

7. Conclusion

In this paper, the origin of the arithmetic codes was described at first. The basic study of arithmetic coding was mainly achieved by J.J. Rissanen[31], G.G. Langdon[32] and R. Pasco[33]. J.J. Rissanen and G.G. Langdon have done the work in IBM, and R. Pasco has done the work in his thesis (Stanford University) and he later joined IBM. They have contributed theoretically and also practically how arithmetic coding/decoding can be utilized.

It can be said that the appearance of arithmetic code for image coding has made a revolution. In Markov-model coding, the extension of sources will be needed in order to keep the coding efficiency, if the probability of the most probable symbol is enough larger than 1/2 and it was not easy to do it by using block code. But by using arithmetic code, the issue was solved in elegant way.

In arithmetic coding, the codeword of occurred symbol sequence can be created independent to other symbol sequences. Also statistic update of the context is possible at any timing. So, if the context and the occurred symbol are given, a universal code design is possible in arithmetic codes.

These two issues can be considered as the main advantage of the arithmetic codes for image coding. Also, a lot of effort have been done to make arithmetic coding more practical. There are trade-off between simplification and efficiency keeping, and further effort will be expected depending on the target sources and applications.

Concerning the comparison of AC and HC, though the codeword creating mechanism is different, the relation of the entropy and mean code length per symbol is theoretically same, if HC is created with alphabetical order as shown in **2.4**. And though codeword is given for the unit of message in HC and for the total symbol sequence in AC, it can be said that there are no theoretical differences in efficiency.

It will be possible to consider AC as the upper concept of HC, or both codes are constructing a kind of duality. There are some reviews considering the AC specifications of specific standards to be general or unique to AC, and I lastly hope such misunderstanding will disappear by letting readers know the concept of AC and HC more deeply.

References

1) P. Elias: Note 1 in Chapter 3 (pp.61-62) of "Information Theory and Coding" by N. Abramson, McGraw-Hill Inc. (1963).
2) F. Ono: "The Entropy Coding of Markov-Model Sources: The necessity of Arithmetic Coding

Operation and its Application to MELCODE", Journal of IIEEJ, Vol.21, No.5, pp. 475-485 (Oct. 1992).

3) E. N. Gilbert, E. F. Moore: "Variable-Length Binary Encodings" BSTJ pp.933-967 (July 1959).
4) F. Ono: "Studies on Arithmetic Codes and their Application to Image Coding", Journal of IIEEJ, Vol.31, No.5, pp. 745-754 (Sept. 2002).
5) T. S. Huang: "Picture Bandwidth Compression", Gordon & Breach (1972) .
6) Y.Yamazaki, Y. Wakahara, H. Teramura: "Digital Facsimile Equipment 'Quick-FAX' Using a New Redundancy Reduction Technique, " NTC '76, 6.2-1 (1976).
7) T. Yamada: "Edge-Difference Coding – A New, Efficient Redundancy Reduction Technique for Facsimile Signals", IEEE Trans. on Comm., Vol.27, Issue 8 (1979).
8) K. Yuki, T. Yamada, Y. Yamazaki, Y. Wakahara: "Relative Element address designate coding scheme for facsimile signals" Journal of IIEEJ, Vol.8, No.4, pp.265-275, Dec. 1979.
9) D. Preuss: "Two Dimensional Facsimile Source Encoding Based on a Markov Model," NTZ, 28, H-10 (1975).
10) R.Ohnishi, Y.Ueno, F.Ono: "Optimization of Facsimile Data Compression" NTC'77 49.1.1-49.1.6 (1977.12).
11) M.Takagi and T.Tsuda, "Band-width compression for facsimile using two-dimensional prediction", Trans. IECE Japan, vol. 56-D, pp. 170-177, Mar. 1973.
12) A. N. Netravali, F. W. Mounts, E. G. Bowen: "Ordering Techniques for Coding of Two-Tone Facsimile Pictures". BSTJ, Vol.55, No.10, Dec.1976.
13) Y.Yasuda, K.Arai: "Facsimile Data Compression by Rearranging Picture Elements", PCS 1977 4-5 (1977)
14) R.Ohnishi, Y.Ueno, F.Ono: "Efficient Coding for Binary Information Sources" Trans. of the IECEJ, Section J, Vol. 60-A, No.12, pp.1114–1121, (1977).
15) F.Ono, Y.Ueno, S.Iwata. R.Ohnishi: "Transmission and Printing of Continuous-tone Images using Multi-level Ordered Dither", NTC'79 53.7 (Dec. 1979)
16) F.Ono, S.Kino. M.Yoshida, T.Kimura : "Bi-level Image coding with MELCODE– Comparison of Block Type Code and Arithmetic type code– ", 7.6.1-7.6.6. Globecom '89 (1989).
17) F. Ono: "The Research of Entropy Coding for Markov-model Images Using Arithmetic Codes", Ph.D Thesis, University of Tokyo, April 1993.
18) F.Ono et al. "Definition of the QM-coder", ISO/IEC JTC 1/SC2/WG8/JBIG N224 R1(1990).
19) G. Langdon, J. Rissanen:" Compression of Black-White Images with Arithmetic Coding", IEEE Trans. on Commun. COM-29, No.6, pp. 858–867 (June 1981)
20) W.B.Pennebaker, J.L.Mitchell, G. G. Langdon Jr., R.B.Arps: "An Overview of the Basic Principles of Q-Coder", IBM Journal of Research and Development, Vol.32, No.6, pp.717–726 (1988).
21) W.B.Pennebaker, J.L.Mitchell: JPEG Still Image Data Compression Standard, Springer New York, NY (1993).
22) D. Taubman, M. Marcellin: JPEG2000 Image Compression Fundamentals, Standards and Practicem, Springer (Nov. 2001).
23) F. Ono, T. Kimura: "A Carry Propagation Control Method with Less Loss Bit for the Arithmetic Coders", Trans on IEICE Vol. J84-A, No.3, pp.414-417 (Mar. 2001) .
24) F. Ono, T. Kimura: "A State Transition Type Probability Estimation Scheme with MPS Counting for Coding Binary Sources", Trans on IEICE Vol.J83-A, No.6, pp. 711-721 (June 2000).
25) D.L. Duttweiler, C. Chamzas: "Probability Estimation in Arithmetic and Adaptive-Huffman Entropy Coders", IEEE Trans. on Image Processing, Vol.4, No.3, pp. 237-246, March 1995.
26) D. Marpe, H. Schwarz, T. Wiegand: "Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard", IEEE Trans. on Circuit and Systems for Video Technology, Vol. 13, No. 7, (JULY 2003).
27) I. Ueno, F. Ono: "Designing Method of State Transition Table-Driven Arithmetic Coder STT-coder", Journal of IIEEJ, Vol.43, No.1, pp. 62-70 (Jan. 2014).
28) I. Ueno, F. Ono: "Probability Estimation and

Dynamic Coding Method for State Transition Table-Driven Arithmetic Coder STT-coder", Journal of IIEEJ, Vol.43, No.1, pp. 71-78 (Jan. 2014).

29) F.Ono: "Basic Verification of Multiplication Free Multi-Alphabet Arithmetic Codes", Annual Conference of IIEEJ, P5-4 (2022).

30) J. J. Rissanen: "Generalized Kraft Inequality and Arithmetic Coding", IBM Journal of Research and Development, Vol.20, Issue 3, May 1976, pp.198–203 (1976).

31) J. Rissanen, G. G. Langdon: "Arithmetic Coding", IBM Journal of Research and Development, Vol. 23, Issue 2, March 1979, pp.149–162 (1979).

32) R. C. Pasco: "Source Coding Algorithm for Fast Data Compression", Ph. D. Thesis, Dept. of Electrical Engineering, Stanford Univ. (May 1976).

Fumitaka ONO (***Honorary Member)***

He has received BE, ME and Ph.D from The University of Tokyo respectively. He has been working with Mitsubishi Electric Corp., and then with Tokyo Polytechnic University. He is currently Professor Emeritus in TPU, and Visiting Researcher of The Univ. of Tokyo. His interested area covers image coding, and image processing. He is IEEE Fellow, IEICE Fellow and IIEEJ Fellow. He has been engaged in the international standardization work since 1985, and had been ISO/IEC JTC 1/SC29/WG1 JBIG Rapporteur. He has received Award from Ministry of Education and Science, Award from Ministry of Industry and Trade, SCAT Chairman's Grand Award, and Contribution Award from IPSJ/ITSCJ for distinguished achievements on image coding and the contribution for standardization activities.

IIEEJ Paper

ECA-Resunet: Lung Segmentation of CT Images Based on an Efficient Channel Attention Mechanism Deep Residual-UNet

Jincheng PENG† *(Student Member)* , Ruigang GE†, Guoyue CHEN‡ , Kazuki SARUTA‡, Yuki TERATA‡

†Akita Prefectural University Graduate School, ‡Akita Prefectural University

<Summary> Precise automatic segmentation of lung organs by computed tomography (CT) is a prerequisite for organ identification, pathological localization and treatment of lung diseases. However, accurate lung segmentation remains a major challenge due to the shape, size and location of the lungs. In recent years, U-net network structures and variations have been applied to various medical image segmentation, but these networks still have limitations and shortcomings in solving the vanishing gradient problem and contextual semantic feature extraction. In this paper, we propose a deep residual network called ECA-Resunet to segment lung organs with an efficient attention mechanism. The structure of ECA-Resunet is similar to Res-UNet. It uses deep residual units to form the entire encoder-decoder network and adds an efficient channel attention mechanism to the encoder. Compared to other networks, the advantage of ECA-Resunet is that it uses a deep residual network, which makes it difficult for the gradient to vanish. The encoder introduces an efficient channel attention mechanism to increase the weight of key regions in order to highlight key features and better learn the semantic features of the image context. This way, we can design a network with fewer settings and achieve better semantic segmentation results without changing the original image size. The results show that evaluation metrics such as Miou Score, Dice Effective Score, and Sensitivity of ECA-Resunet outperform those of the comparison network.

Keywords: lung segmentation, Res-UNet, efficient channel attention network, computed tomography image

1. Introduction

In 2019, the COVID-19 swept the world. A quick and accurate diagnosis of patients is an important part of preventing COVID-19. Besides regular RT-PCR detection, analyzing lung CT images is also very important for the treatment and diagnosis of the COVID-19[1)]. Therefore, a high-precision lung organ segmentation method is needed. At the same time, the doctor can diagnose a few lung diseases consisting of tuberculosis, pneumonia, and lung cancer. with the assist of accurate segmentation images of lung organs to diagnose and plan surgical treatment for a few lung diseases. However, due to the complexity of the anatomy of lung organs, the low contrast of adjacent organs, the low incidence, shape and size of lung tissues and organs vary from individual to individual, manual segmentation is vulnerable to errors.

With the development of modern tomography modalities such as computed tomography (CT) and magnetic resonance imaging (MRI), computer image-guided diagnosis of lung diseases has been used for regular diagnosis and monitoring of lung organ diseases. High-resolution images facilitate visualization of the entire lung organ concerning surrounding anatomy. Traditional image segmentation techniques, such as threshold-based segmentation, edge detection segmentation, and region-based segmentation, can replace manual segmentation[2)-4)], and rough subdivide the target organs and background within the image after digitizing the image. However, these methods are inefficient and low segmentation fine-grained, there is an urgent need for automatic and accurate lung organ segmentation methods.

In recent years, the application of convolutional neural networks in image segmentation has made extraordinary progress. The literature[5)-7)] has used various networks of convolutional neurons to improve the precision of the semantic segmentation of images with good results. Meanwhile, the literature[8)-10)] has applied convolutional neural networks to the segmentation aspects of medical-imagery.

In particular, the encoder-decoder network structure of U-net and its variants has achieved state-of-the-art performance in various medical image segmentation tasks. The U-net is a convoluted network image segmentation model with a U-shaped structure for biomedical applications. The network pioneers a two-part encoder and decoder network structure. However, the network effect on the location and boundary segmentation of organs is not satisfactory, Achieved mean IoU of 0.921 on the 2015 ISBI Cell Challenge dataset. The literature 11) prooposed a deeper

Res-UNet convolutional network plus random conditional fields for segmentation of skin cancer regions. It learns from the ResNet network and replaces the normal convolutional blocks in U-net with the residual convolutional blocks in the network encoder. It can solve the problem that vanishing gradients with increasing depth. However, semantic segmentation is not fully equivalent to pixel-wise prediction and pixel grouping. Contextual information between these network image pixel channels connection is not spotlighted.

As researchers have studied attention mechanisms to obtain contextual information about image pixel channels connection by attention mechanism. Reference 12) proposes a squeeze and excitation network. By learning channel attention of each convolutional block, it can increase the information weight of key regions to highlight the features of target regions, bringing significant performance improvements to CNN network architecture of different depths. Based on SE-Net, Reference 13) proposes a lightweight network strategy for local cross-channel interaction without dimensionality reduction and adaptive selection of one-dimensional convolution kernel size, that significantly reduces the complexity of the SE-Net model while maintaining efficiency. The literature 14) proposed a Dual-Attention and Skip Connection to combine the feature maps of the encoder and decoder to complete pixel-level rib fracture detection. The method has attained a sensitivity 89.58%, and the average FROC score is 81.28%. Reference 15) proposes using U-Net with attention mechanism for abdominal organ segmentation, that improves the sensitivity of foreground pixels useful for specific tasks and obtains satisfactory results. Scored 84.08% CSD on the pancreas dataset. From the network experimental analysis of Attention U-Net network in the deep network convergence, and the acquisition of contextual information is still slightly inadequate.

Based on the above networks and research on the attention mechanism of ECA networks, this paper proposes an ECA-Resunet networks: a deep residual encoder-decoder network architecture with an efficient channel attention mechanism for CT images of lung organs segmentation[16]. Since medical CT segmentation requires more precision segmentation, not just simple pixel classification. To define these target lung organ boundaries, we must emphasize the related pixels to improve the segmentation accuracy of key regions. We used an Efficient Channel Attention mechanism. The Efficient Channel Attention module proposes a local cross-channel interaction strategy without dimensionality reduction and a lightweight attention mechanism method for adaptively selecting the size of 1D convolution kernels. the attention mechanism is used to learn high-level semantic features, increase the information weight of key regions to emphasize the properties of the target region, suppress the interference of irrelevant regions in the image, and model the feature mapping dependencies between different channels. Specifically, ECA-Resunet is similar to Res-UNet network architecture that includes an encoder, a decoder, and a bridge-like layer skip connection. The encoder of ECA-Resunet contains an efficient channel attention module. Meanwhile encoder and decoder adopted the deep residual structure improves the utilization of image features by passing depth gradient information back to shallow layers while avoiding gradient vanishing and network degradation. Enables the network to obtain high accuracy while increasing the robustness and stability of the network.

We trained our proposed network structure method on the 2016 Lung Node Analysis (LUNA) competition dataset. 80% of the competitive Lung Node Analysis (LUNA) dataset was used as the training dataset and 20% of the data was used as the validation set. The result shows that the mean IOU, Dice coefficient (DC) and sensitivity score of ECA-Resunet in the lung organ validation set were 93.84%, 96.82% and 94.63% respectively. The boundary of lung organ segmentation results is more smooth and accurate. The network addresses the stability of automatic segmentation and accuracy of segmentation results, and future studies will investigate whether it can be used for clinical CT segmentation.

2. Related Work

2.1 Network structure

The entire network structure is shown in **Fig. 1**. The Res-UNet with improved encoder and decoder structure is adopted as the backbone architecture of the network. Res-UNet consists of three parts: encoding, bridge and decoding[17]. The encoder and decoder form a symmetrical U shape. The encoder performs a downsampling process on the input lung computed tomography image. High-level semantic information in the image is extracted layer by layer, and a small feature tensor is obtained. Before each deep residual module, we add an ECA (Effective Channel Attention) module to improve the effectiveness of feature map ping. The middle part serves as a bridge, connecting the encoding and decoding paths that connect the encoder and the decoder. The decoder on the right also consists of a deep residual module that performs 3 layers upsampling on the

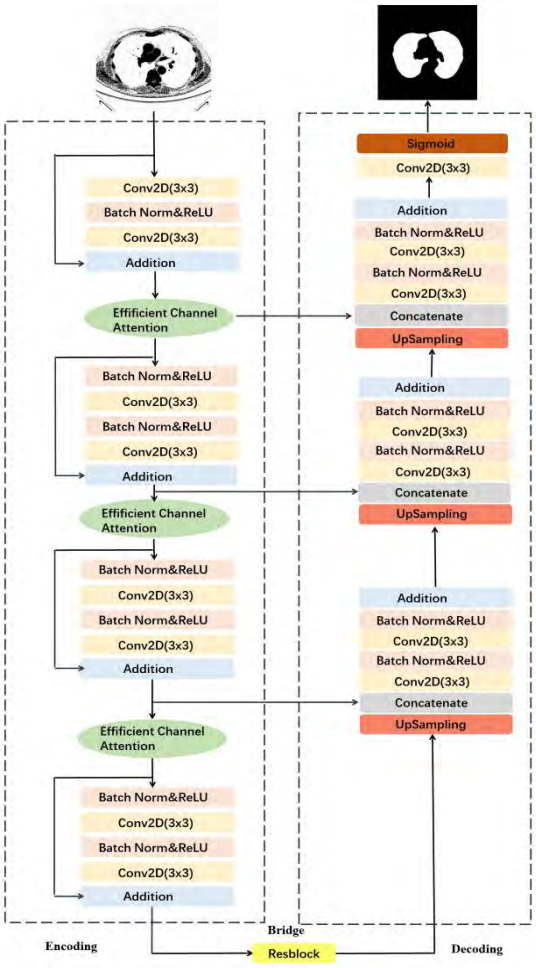


Fig. 1 The architecture of the proposed ECA-Resunet

feature tensor. Each layer of the decoder predicts a segmentation map for calculating the loss function, and recovers the extracted features into a pixel-level classification. In this way, the nearest neighbor up a sampling of the feature-map can be obtained from the lower level and concatenated with feature mappings obtained from their corresponding encoding paths, finally obtaining a binary image with the same size as the original image.

2.2 Residual module block

CNN networks improve the performance of neural networks with increasing network depth. However, increasing the network can hinder training at the same time, and the network may suffer from degradation problems. To overcome these problems, literature 18) proposes a deep residual neural network to facilitate training and solve the problem of network degradation. The residual module design in ECA-Resunet is shown in **Fig. 2**, which consists of a BN layer, Relu activation layer and 3×3 convolutional layers form a

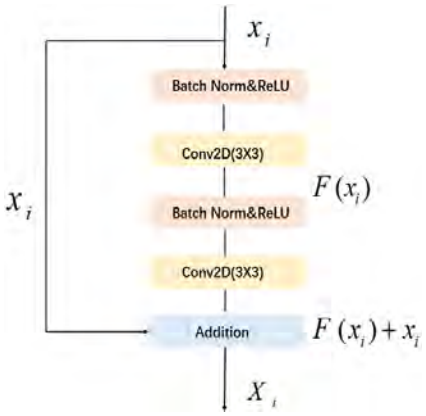


Fig. 2 Residual module with identity mapping used in the proposed ECA-Resunet

residual group, and each residual module contains two stacked residual groups. Each residual unit is connected through residuals between layers of different depths, Instead of a direct continuous long connection. The remaining blocks can distribute information on each layer to build deeper neural networks. The advantages of this structure are: 1) The skip connection with the residual units and between the low-level and high-level networks will facilitate information propagation without degradation, avoiding the problem of gradient dispersion in the encoder caused by deepening the network. 2) At the same time, the interdependence of the channels is improved, and the speed of training convergence is accelerated. The residual unit can simplify the training of the network, reduce the overall parameters of the neural network, reduce the computational cost, and design a network with fewer parameters but better semantic segmentation performance.

2.3 Efficient Channel Attention module block

We begin with a study of the SE-Net network. SE-Net filters the features of different channel dimensions through global average clustering, and then uses the model to filter the features and select important features for improvement. The SE module first uses global mean pooling for each channel separately, and then uses two nonlinear fully connected layers, finally using the sigmoid function to generate the channel weights. The two fully connected layers are designed to capture nonlinear interactions between channels. In order to control the complexity of the overall feature, the dimension of the feature vector is reduced. However, the dimensionality reduction process reduces the detail of the feature description, so this operation has a side effect on the channel attention prediction, thus reducing the accuracy of the network.

Effective Channel Attention Network (ECA-Net) is an improved channel attention module based on the SE-Net. Through a local cross-channel interaction strategy without dimensionality reduction and a method of adaptively selecting the size of one-dimensional convolution kernels, the performance of complex attention modules can be improved based on increasing the minimum parameters. At the same time, reduces the complexity of the dimensionality reduction operation of the SE-Net model. **Figure 3** shows how the attention mechanism of the ECA channel work. The original input image features are shown on the left. All features without dimensionality reduction are obtained by global mean pooling (GAP); local cross-channel interaction information is collected by considering each channel and its K neighbors. Here, instead of using a fully connected layer in the SE module, one-dimensional convolution is used to capture local cross-channel interaction information. One-dimensional convolution includes hyper parameters K, i.e. the size of the convolution kernel, It represents the coverage of local cross-channel interactions. Since we need large-scale convolution to analyze the contour images of the whole lung organ, the value of K is 5 in this experiment. Then use the sigmoid function to generate a weight score for each channel, and then the original input feature is combined with the channel weight to obtain the feature with channel attention. The channel attention can be computed as:

$$\begin{aligned} \mathrm{CA}(x) &= \delta(f_{conv}(\mathrm{AvePool}(x))) \\ &= \delta(W(\mathrm{AvePool}(x))) \end{aligned} \tag{1}$$

where AvePool represents the global average pooling operation, W is the weight of the f_{conv} convolution layer, and is the sigmoid of the activation function. The network constructed with this module is easier to extract the salient features of the image based on channel dimensions, capture interaction information between channels, increase the weight of key regions, and obtain more abundant feature information of key regions.

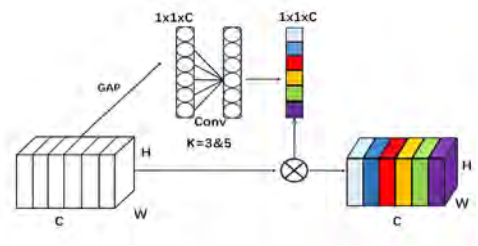


Fig. 3 Diagram of efficient channel attention (ECA) module

2.4 Loss function

The loss function is used to estimate the inconsistency between the predicted segmentation map f(x) of the model and the true background value y. In the actual segmentation of lung organs, due to the unbalanced proportions of foreground (lung) and background (non-lung), lung organs are likely to occupy a small area in the image, resulting in a decrease in segmentation accuracy. Therefore, using only the binary cross-entropy loss function (BCE loss) is not optimal for this type of task. This paper uses a combined loss function of the Dice coefficient loss function plus binary cross-entropy loss function to predict the similarity between the lung organ segmentation region and the labeled true region, instead of the usual binary cross-entropy loss[19]. The combined loss function is an overlap metric and is often used to evaluate the segmentation performance. Even if the area of the lung organ is smaller than the background area, there is no effect on the Dice value. Many experiments have confirmed that the Dice loss function can solve the problem of class imbalance. The Dice coefficient loss function and the two Combining the loss function using the Dice coefficient loss function plus the binary cross-entropy loss function improves the training. The Dice coefficient loss function is calculated as follows:

$$L_{Dice} = 1 - \frac{2\sum_i^N p(k,i)g(k,i)}{k\sum_i^N p_{(k,i)}^2 + \sum_i^N g_{(k,i)}^2} \tag{2}$$

Where N is the number of pixels, and $p(k,i) \in [0,1]$ and $g(k,i)$ represent the prediction probability and Truth label of class k , it takes 0 or 1. k is the class number. The final loss f unction is defined as:

$$L_{loss} = L_{bce} + L_{Dice} \tag{3}$$

2.5 Evaluation method

In this paper, the most commonly used medical image evaluation metrics, such as Dice coefficient (DC), Mean IoU and sensitivity, are used to evaluate the training data. The Dice Coefficient (DC) takes care of the class imbalance problem and hence it has been used as the most common evaluation metric. The Dice coefficient represents the similarity between the output (y) and the label (x) of the model's prediction. The higher the lung organ similarity and label confidence value, the higher the Dice coefficient, and the better the segmentation effect. The calculation formula of the Dice coefficient is as follows:

The problem of binary semantic segmentation of images is essentially the problem of binary classification of pixels. We consider pixels in lung CT slices to be positive samples, background pixels to be negative samples, TP and TN indicate that the pixel is the predicted category matches the actual category, and FP indicates that the pixel is background but is predicted to be a lung organ, indicating that the pixel is incorrectly classified as a false positive. and FN indicates that the pixel is a lung organ but is predicted to be background. This indicates that the pixel is incorrectly classified as a false negative.

The mean IOU average intersection union ratio is one of the most commonly used evaluation metrics of semantic segmentation[20]. The ratio of intersection and concatenation of the predicted and background true values for each class of pixel segmentation in the lung organ CT slice is calculated and then averaged. Using the above definition, the calculation formula is as follows:

$$MeanIoU = \frac{1}{k}\sum_{i=0}^{k}\frac{TP}{FP+TP+FN} \tag{4}$$

where k is the division category.

Sensitivity is sensibility, which represents the ratio of pixels correctly predicted as lung organ category to background real value pixels. The range of sensitivity coefficient is [0,1]. A value closer to 1 indicates that a higher percentage of lung organ pixels are correctly classified, i.e. the higher the accuracy of the model. The sensitivity is expressed as follows:

$$S = \frac{TP}{TP+FN} \tag{5}$$

3. Experiments

3.1 Database

We use the 2016 Lung Nodule Analysis (LUNA) dataset to evaluate our model. This is the segmentation of lung organ tissue structures in 2D CT images as part of Lung Nodule Analysis (LUNA) challenge[21]. The LUNA16 dataset contains labeled low-dose lung CT image data from 888 patients. For each patient, the CT scan data consists of a variable number of images (typically around 100-400, each image is an axial slice). Primary validation of nodule detection and reducing false positives. Since full segmentation of lung organs is a further basis for lung nodule detection, The original data image is a 3D image, which is composed of a series of 2D images with a slice interval of 2 mm. As this paper is intended to perform validation of 2D lung contours,

$$DiceCofficient(DC) = \frac{2\times|X\cap Y|}{|X|+|Y|} \tag{6}$$

in order to facilitate validation, 534 2D samples without lung nodules were randomly sliced from the original dataset, and these samples has been threshold processed and threshold normalized to highlight the features of the lung organs to obtain more accurate segmentation of the true labels. We use this challenge dataset to evaluate the structure of our proposed ECA-Resunet model. The dataset contains 534 2D samples and corresponding true labels, and the data size is 512×512. Since the image is too large and wastes storage space, the data is cropped to 256×256. The processed data were divided into 427 training datasets and 107 validation datasets.

3.2 Experimental process

All training and testing experiments are performed onwork stations. The CPU used is Intel (R) Xeon (R) CPU e5-2630V4@2.20 GHz. The GPU uses two NVIDI-Aforcedx1080ti with 12GB GPU memory, with a total memory of 24G. All network architectures are implemented with pytorch 1.7.1+cu101 framework. We start the training with a batch size of 5. The proposed architecture is optimized by the Adam optimizer with the learning rate initially set to 0.001. The loss function is used as the combined loss function mentioned earlier.

3.3 Comparisons with state-of-the-arts

Table 1 lists the experimental results obtained from several state-of-the-art segmentation networks. The table below shows that ECA-Resunet outperforms U-Net and Attention-UNet on evaluation metrics such as average IOU, dice efficiency (DC), and sensitivity, and the parameters are not significantly improved compared to U-Net. The results show that the values of the ECA-Resunet network and CE-Net are relatively close.

CE-Net is the current mainstream segmentation network, using dense atrous convolution module (DAC) and a residual multi-pooling module similar to a pyramid structure[22].

Table 1 The evaluation results of all the models on Lung Nodule Analysis (LUNA) dataset

NET	MIou	DC	Sensitivity	Parameter
U-Net	0.8935	0.9432	0.9274	5.88M
Attention-UNet	0.9337	0.9657	0.9378	35M
CE-Net	0.9350	0.9664	0.9354	38M
ECA-Resunet	**0.9384**	**0.9682**	**0.9463**	**12.73M**

This design will increase the number of parameters, increase the amount of computation, and increase the training and testing time.

Figure 4 shows the comparison between ECA-Resunet and the training set Dice coefficient (DC) of the current more popular medical image segmentation algorithm. **Figure 5** shows the evaluation results of the designed ECA summary on foreground (lung) and background(non-lung) segmentation of CT images of clinical lung organs, and compared with the current popular medical image segmentation algorithm U-Net, CE-Net, Attention-UNet. From left to right, each column contains the original image, true labels, and proposed method test results. As can be seen from the above figure, ECA-Resunet improves segmentation performance by focusing on related key regions. Compared with other methods, ECA-Resunet reduces the area of prediction error. The edge tissue segmentation of lung organs is smoother. It proves that the ECA-Resunet result best.

3.4 Ablation study

In this section, we use ablation experiments to evaluate the various components of the proposed ECA-Resunet network model structure, and verified the effectiveness of the effective channel attention mechanism and the deep residual structure adopted by the proposed network, U-Net, U-Net +SE module, Unet+ECA module, Res-UNet module, Res-UNet+ECA module are compared respectively. The details are shown in **Table 2**. As can be seen from the table, the single SE-Unet and ECA-Unet can be seen from the Table 2 to be much better than Unet on MIou and DC. Due to the lightweight attention mechanism, In the previous narrative, the ECA attention mechanism uses a local cross-channel interaction strategy without dimensionality reduction and adaptive selection of the one-dimensional convolutional kernel size, because there is a loss of semantic details in the process of dimensionality reduction, so the use of ECA-Unet network is better than SE-Unet network. However, compared with Res-UNet, ECA-Resunet does not increase the number of parameters, and the MIou of the partition increases by 1.5%. Compared with the U-net network, the MIou increases by 4.5%, thereby obtaining higher results. Therefore, despite the increase in U-net network computation costs, these network computational costs are necessary. This shows that the attention mechanism significantly improves the network performance after the introduction of the ECA module, and also has a certain effect on the overall model segmentation.

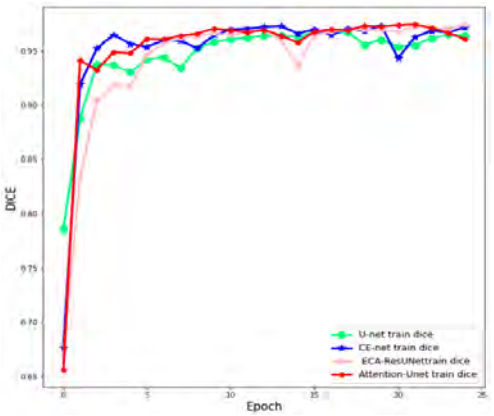


Fig. 4 Different model training dice on the Lung Nodule Analysis (LUNA) dataset

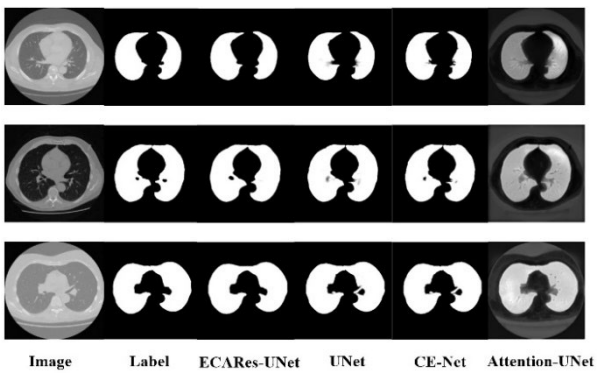


Fig. 5 Different segmentation model segmentation effect

Table 2 The evaluation results of ablation study on Lung Nodule Analysis (LUNA) dataset

NET	MIou	DC	Sensitivity	Parameter
U-Net	0.8935	0.9432	0.9274	5.88M
SE-Unet	0.9270	0.9621	0.9313	7.5MB
ECA-Unet	**0.9306**	**0.9640**	**0.9357**	**7.5MB**
Res-UNet	0.9235	0.9602	0.9279	12.73MB
ECA-Resunet	**0.9384**	**0.9682**	**0.9463**	**12.73MB**

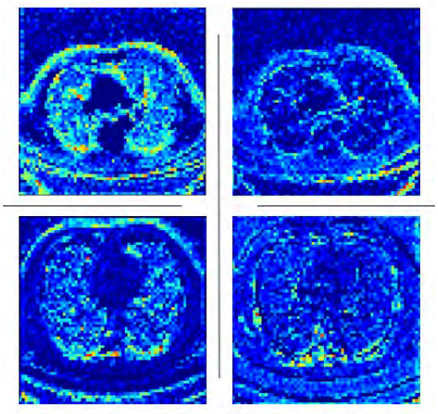


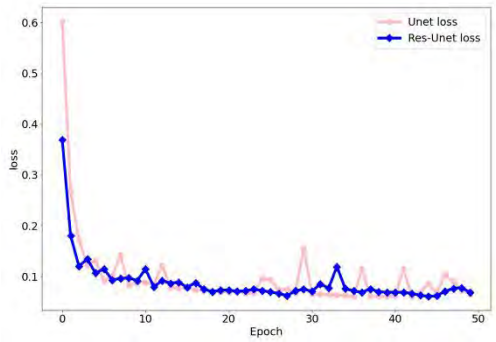
Fig. 6 (a) Activation information of the ECA-Resunet network model feature map (b) Activation information of the U-net network model feature map

To verify the capture of contextual information between image channels by the ECA attention mechanism module. We extracted separately the feature map from the third layer downsampling on a channel of two lung organ images in the dataset. From the comparison of the ECA-Resunet network model in **Fig. 6(a)** on the left and the U-net model in Fig. 6(b) on the right, it can be seen that the ECA-Resunet network model has more colored activation information in the feature maps. In this way, it is shown that the ECA-Resunet network model is better than the U-net model in capturing semantic information, and there is fine-grained feature map with less misfortune in semantic details.

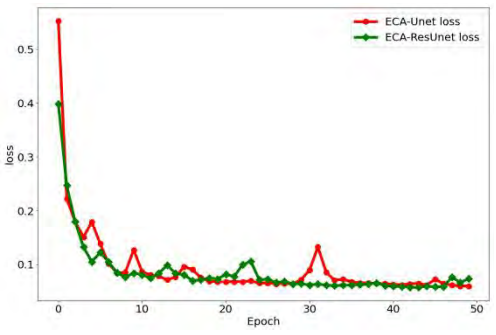
From the loss function curves of **Fig. 7(a)** and Fig. 7(b), the fluctuation in the network validation loss assembly after adding the residual structure is low, but the loss converts quicker, more stable and smoother than the structure without residual modules. It is proved that the deep residual model not only can performance be improved but also can effectively overcome the problem of gradient vanishing/exploding in the initial stage of training.

4. Conclusion

In this paper, we aim to enable the network to better learn more efficient features and obtain more accurate lung organ segmentation results. We start with the U-shaped encoder-decoder network architecture, In the network encoder, the normal convolutional blocks in U-Net are replaced by residual convolutional blocks, which has the advantage of solving the problem of disappearing deep gradients as the depth of the network increases. But only in this way contextual information between image channels is not improved.



(a) Validation loss for U-net and Res-UNet



(b) Validation loss for ECA-Unet and ECA-Resunet

Fig. 7 (a) Validation loss for U-net and Res-UNet on Lung Nodule Analysis (LUNA) dataset (b) Validation loss for ECA-Unet and ECA-Resunet on Lung Nodule Analysis (LUNA) dataset

So we are leveraging the strategy to improve the compression SE-Net network through effective channel attention. Use a strategy for local cross-channel interaction between channels with no reduction in dimensionality. A lightweight attention mechanism method with adaptive selection of one-dimensional convolution kernel size. The attention mechanism can emphasize the characteristics of the target area by increasing the information weight of the key regions, suppress the interference of irrelevant areas in the image, and because the attention mechanism is a lightweight network structure, it will not increase too many network parameters and will not change the resolution of the original image. Dramatically reduce the complexity of the original extrusion-excitation design while maintaining performance. This enhanced strategy acquires channel focus to improve context information between image channels. With both of the above enhancements to the U-shaped encoder-decoder network.

We design a deep residual encoder-decoder called ECA-Resunet Effective Channel Attention Mechanism server network architecture. It maintains the overall U-shaped network of encoders-decoders. By adding a residual module and an effective attention module, solve the problem of capturing contextual information of image channels and gradient vanishing explosion by adding a residual module and an effective attention module. A Dice coefficient (DC) score of 96.82% was obtained in the 2016 Lung Nodal Analysis (LUNA) Challenge lung organ segmentation data. In addition, the network surpasses other similar methods for evaluation parameters, and with significant contextual semantic feature extraction, more accurate and smoother lung organ contour segmentation, and less fine-grained detail loss. In addition, because the network parameters do not increase significantly, unnecessary network overhead is reduced, and perfect lung organ segmentation is achieved while ensuring accuracy, which can better assist doctors in lung organ diagnosis and surgery planning. Future improvements can be made to this structure by incorporating the relationship between long distance positions of pixels. Other enhancements to the network framework for semantic information capture.

References

1) L. Li, L. Qin, Z. Xu, et al: “Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT”, Radiology (2020).
2) Y. E. Erdi, O. Mawlawi, S. M. Larson, M. Imbriaco, H. Yeung, R. Finn: “Segmentation of Lung Lesion Volume by Adaptive Positron Emission Tomography Image Thresholding”, Cancer: Interdisciplinary International Journal of the American Cancer Society, Vol.80, No.12, pp. 2505–2509 (1997)
3) M. N. Saad, Z. Muda, N. S. Ashaari, H. A. Hamid: “Image Segmentation for Lung Region in Chest X-ray Images Using Edge Detection and Morphology” 2014 IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2014), pp.46–51 (2014).
4) P. Annangi, S. Thiruvenkadam, A. Raja, H. Xu, X. W. Sun, L. Mao: “A Region Based Active Contour Method for X-ray Lung Segmentation Using Prior Shape and Low Level Features”, IEEE International symposium on biomedical imaging: from nano to macro, pp.892–895 (2010).
5) J. Long, E. Shelhamer, T. Darrell: “Fully Convolutional Networks for Semantic Segmentation”, Proc. of the IEEE conference on computer vision and pattern recognition, pp.3431–-3440 (2015).
6) V. Badrinarayanan, A. Kendall, R. Cipolla: “Segnet: A Deep Convolutional Encoder-decoder Architecture for Image Segmentation”, IEEE Trans. on pattern analysis and machine intelligence, Vol.39, No.12, pp. 2481–2495 (2017).
7) L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille: “Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs”, IEEE Trans. on pattern analysis and machine intelligence, Vol.40, No.4, pp. 834–848 (2017).
8) S. Miao, S. Piat, P. Fischer, A. Tuysuzoglu, P. Mewes, T. Mansi, R.Liao: “Dilated FCN for Multi-agent 2D/3D Medical Image Registration”, Proc. of the AAAI Conference on Artificial Intelligence, Vol.32, No.1 (2018).
9) C. S. Chang, J. F. Lin, M. C. Lee, C. Palm: “Semantic Lung Segmentation Using Convolutional Neural Networks”, Bildverarbeitung fur die Medizin 2020. Springer Vieweg, Wiesbaden, pp.75–80 (2020).
10) O. Ronneberger, P. Fischer, T. Brox: “U-net: Convolutional Networks for Biomedical Image Segmentation”, International Conference on Medical Image Computing and Computer-assisted Intervention. Springer, Cham, pp.234–241 (2015).
11) G. M. Venkatesh, Y. G. Naresh, S. Little, N. E. O’Connor: “A Deep Residual Architecture for Skin Lesion Segmentation”, OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis. Springer, Cham, pp.277–284 (2018).
12) J. Hu, L. Shen, G. Sun: “Squeeze-and-excitation Networks”, Proc. of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141 (2018).
13) Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu: “ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks” , Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020).
14) Z. Zhou, Z. Fu, J. Jia, J. Lv: “Rib Fracture Detection with Dual-Attention Enhanced U-Net”, Computational and Mathematical Methods in Medicine, (2022).
15) O. Oktay, J. Schlemper, L. L. Folgoc, et al: “Attention U-net:Learning Where to Look for the Pancreas”, arXiv preprint arXiv, pp.1804.03999 (2018).
16) J. Wen, Z. Li, Z. Shen, Y. Zheng, S. Zheng: “Squeeze-and-Excitation Encoder-Decoder Network for Kidney and Kidney Tumor Segmentation in CT Images”, International Challenge on Kidney and Kidney Tumor Segmentation. Springer, Cham, pp.71–79 (2022).
17) Y. Dong, H. Feng, Z. Xu, Y. T. Chen, Q. Li: “Attention Res-Unet: an Efficient Shadow Detection Algorithm”, Journal of ZheJiang University (Engineering Science), Vol.53, No.2, pp.373–81 (2019).
18) K. He, X. Zhang, S. Ren, J. Sun: “Deep Residual Learning for Image Recognition”, Proc. of the IEEE conference on computer vision and pattern recognition, pp.770–778 (2016).
19) W. R. Crum, O. Camara, D. L. Hill: “Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis”, IEEE Trans. on medical imaging, Vol.25, No.11, pp.1451–1461 (2006).
20) M. A. Rahman, Y. Wang: “Optimizing Intersection-over-union in Deep Neural Networks for Image Segmentation”, International symposium on visual computing. Springer, Cham, pp.234–244 (2016).
21) A. A. A. Setio, A. Traverso, T. D. Bel, et al: “Validation, Comparison, and Combination of Algorithms for Automatic Detection of Pulmonary Nodules in Computed Tomography Images: the LUNA16 Challenge”, Medical image analysis, pp.1–13 (2017).
22) Z. Gu, J. Cheng, H. Fu, et al: “Ce-net: Context Encoder Network for 2D Medical Image Segmentation”, IEEE Transactions on medical imaging, Vol.38, No.10, pp.2281–2292 (2019).

(Received July 26, 2022)
(Revised February 6, 2023)

Jincheng PENG *(Student Member)*

He received the ME degree from Chengdu University Of Technology in 2020. From 2021, he has been staying at PhD. course in Systems Science and Technology at Akita Prefectural University Graduate School. His research interests include medical image segmentation, medical image processing, and deep learning.

Ruigang GE

He received his ME degree from Ashikaga University, Japan, in 2019 and is currently pursuing a PhD. in Systems Science and Technology at Akita Prefectural University Graduate School. His research focuses on visual recognition, deep model ensemble, medical image classification and segmentation, as well as the application of traditional image techniques to medical imaging.

Guoyue CHEN

He received his ME and PhD degrees from Tohoku University, Japan in 1993 and 1996, respectively. Then, he was an assistant professor in Tohoku University, Japan. He is currently a professor in Akita Prefecture University, Japan. His research interests include digital signal processing and its applications to active noise control, image processing and medical image processing.

Kazuki SARUTA

He received his BE and ME degrees from Akita University, Japan, in 1991 and 1993, and his PhD degree from Tohoku University, Japan, in 1996. Then, he was a lecturer in Yamagata University. He is currently a professor in Akita Prefecture University, Japan. He has been engaged in research on pattern recognition and image processing.

Yuki TERATA

He received his BE degree from Akita University, Japan, in 1998, and the ME and PhD degrees from the same university in 2000 and 2008. He is an associate professor in Akita Prefecture University, Japan. He is interested in VR/AR for cognitive psychology of elderly people.

Call for Papers

The 8th IIEEJ International Conference on Image Electronics and Visual Computing 2024 (IEVC2024)

National Cheng Kung University, Tainan City, Taiwan / Mar. 11-14, 2024

https://www.iieej.org/en/ievc2024/

Purpose:

The International Conference on Image Electronics and Visual Computing 2024 (IEVC2024) is scheduled in Tainan City, Taiwan, on March 11-14, 2024, as the 8th international academic event of Image Electronics Engineers of Japan (IIEEJ). Past IEVC conferences were held in Cairns, Australia (2007); Nice, France (2010); Kuching, Malaysia (2012); Koh Samui, Thailand (2014); Danang, Vietnam (2017); Bali, Indonesia (2019); and online (2021). The conference aims to bring together researchers, engineers, developers, and students from various fields in academia and industry to discuss the latest research, standards, developments, implementations, and application systems in all image electronics and visual computing areas.

Topics:

The conference will cover a broad set of research topics including, but not limited to, the following:

- ✧ 3D image processing
- ✧ Bioinformatics and authentication
- ✧ Computer vision
- ✧ Data hiding
- ✧ Image analysis and recognition
- ✧ Image and video coding
- ✧ Image and video retrieval
- ✧ Image assessment
- ✧ Image restoration
- ✧ Mobile image communication
- ✧ Motion analysis
- ✧ Object detection
- ✧ Printing and display technologies
- ✧ Segmentation and classification
- ✧ Smart display
- ✧ Versatile media appliance
- ✧ Animation
- ✧ Content production
- ✧ Extended Reality
- ✧ Metaverse
- ✧ Modeling
- ✧ Non-photorealistic rendering
- ✧ Rendering
- ✧ Visual computing
- ✧ Visualization
- ✧ Architectural industry mondiale
- ✧ Artificial intelligence and deep learning
- ✧ Big data and cloud computing
- ✧ Content delivery network
- ✧ Digital museum, digital archiving
- ✧ Generative AI
- ✧ Hardware and software implementation
- ✧ Interaction
- ✧ International standards
- ✧ Security and privacy
- ✧ Social secured cybertechnology
- ✧ Unmanned Aerial Vehicle
- ✧ Visual and hearing impaired support
- ✧ Visual communication

Paper submission:

The official language is English and authors should submit their papers as PDF through the online submission system, which will be available around Jun. 2023 at the following IEVC2024 official website:

https://www.iieej.org/en/ievc2024/ .

The paper submission guide and IEVC formats (TeX format/MS Word format) will be also provided at this site. The organizing committee particularly encourages graduate students to present their works in the special sessions that are now planned by the committee of the conference.

General Papers:

The general papers category is divided into two types: Journal track and Conference track.

Journal Track:

Journal track aims to publish the papers on the journal in addition to the publishing in the conference, with a quick review process. This type of papers will appear in a special issue on "Journal Track Papers in IEVC2024" in the IIEEJ Transactions on Image Electronics and Visual Computing, Vol. 12, No. 2 (June, 2024), if accepted through journal review process. The authors have to prepare two types of papers different in the amount, the paper for the conference and the paper for the journal. The latter one is the extended version of the former one. **Note that paper for the journal should follow the "guidance for paper submission" available from the web-site of IIEEJ, to be finally published in the IIEEJ Transactions.**

Important Dates

- Pre-Entry Submission (title, authors, 100 words abstract):	Sep. 15, Friday, 2023
- Paper Submission (2-4 pages for the conference)	Sep. 29, Friday, 2023
- Paper Submission (8 pages for the journal):	Oct. 27, Friday, 2023
- Notification of Conference Acceptance:	Nov. 17, Friday, 2023
- Camera-Ready Paper (2-4 pages for the conference):	Dec. 15, Friday, 2023

Conference Track:

Conference track aims to present the papers about recent results and preliminary work at IEVC2024. The authors are required to submit a paper of which length is 2-4 pages. Accepted papers will be published in online proceedings of IEVC2024 (indexed by J-stage) and in the USB proceedings. Rejected papers in the conference track can be resubmitted as late breaking papers **according to the authors' requests in advance**.

Important Dates

- Pre-Entry Submission (title, authors, 100 words abstract):	Sep. 15, Friday, 2023
- Paper Submission (2-4 pages):	Sep. 29, Friday, 2023
- Notification of Acceptance:	Nov. 17, Friday, 2023
- Camera-Ready Paper (2-4 pages):	Dec. 15, Friday, 2023

Late Breaking Papers:

All suitably submitted papers for this category will be accepted for the conference. The authors are required to submit an abstract of which length is 1-2 pages, and select one from the following two types: 1) Technical papers or 2) Art/Demo papers. All the registered papers as late breaking papers will be published only in the USB proceedings of IEVC2024.

Important Dates

- Pre-Entry Submission (title, authors):	Nov. 17, Friday, 2023
- Abstract Submission (1-2 pages):	Nov. 24, Friday, 2023
- Notification of Acceptance:	Dec. 8, Friday, 2023
- Camera-Ready Paper (1-2 pages):	Dec. 15, Friday, 2023

Further information:

After the conference, the Trans. on IEVC of IIEEJ is planning a forthcoming special issue on "Extended Papers Presented in IEVC2024", which will be published in December 2024. More detailed information will be notified on the IEVC2024 web-site and the Journal of IIEEJ.

Call for Papers

Special Issue on
Design and Implementation Technologies to Support Immersive Video Communication and Distribution

IIEEJ Editorial Committee

In recent years, there has been rapid progress in advanced image processing and high-definition video services, driven by advancements in semiconductor technology and information communication technology. Specifically, in addition to improving spatial resolution, frame rates, dynamic range, and color gamut, and other aspects of image quality, there has been an increase in the diversity of immersive media, such as 360-degree video, VR/AR/MR, and so on. Research on video compression, communication, and distribution technologies related to these developments is also thriving. Furthermore, device technologies are essential for the communication and distribution of these videos. In order to make the most of the benefits of further progress in higher integration and lower power consumption in the device technology, there is a strong demand for further progress in hardware/software design technology and implementation technology, such as how to design and implement the required functions.

In this special issue, we invite various categories (Ordinary paper, Short paper, System development paper, Practice Oriented Paper) of papers on hardware/software design technology and implementation technology that support communication and distribution of immersive, high-definition video. We look forward to receiving your contributions.

1. Topics covered include but not limited to

   Video Communication, Video Distribution, Hardware/Software Design Technology, LSI and System, Image Compression, High-definition Video, Immersive Media, 360-degree Video, VR/AR/MR, 3D Related Design Technologies

2. Treatment of papers

   Submission paper style format and double-blind peer review process are the same as the regular paper. If the number of accepted papers is less than the minimum number for the special issue, the acceptance paper will be published as the regular contributed paper. We ask for your understanding and cooperation.

3. Publication of Special Issue:

   IIEEJ Transactions on Image Electronics and Visual Computing Vol.12, No.1 (June 2024)

4. Submission Deadline:

   Tuesday, October 31, 2023

5. Contact details for Inquiries:

   IIEEJ Office E-mail: hensyu@iieej.org

6. Online Submission URL: http://www.editorialmanager.com/iieej/

Revised: January 6, 2017

Revised: July 6, 2018

Guidance for Paper Submission

1. Submission of Papers

(1) Preparation before submission

- The authors should download "Guidance for Paper Submission" and "Style Format" from the "Academic Journals", "English Journals" section of the Society website and prepare the paper for submission.
- Two versions of "Style Format" are available, TeX and MS Word. To reduce publishing costs and effort, use of TeX version is recommended.
- There are four categories of manuscripts as follows:
  - Ordinary paper: It should be a scholarly thesis on a unique study, development or investigation concerning image electronics engineering. This is an ordinary paper to propose new ideas and will be evaluated for novelty, utility, reliability and comprehensibility. As a general rule, the authors are requested to summarize a paper within eight pages.
  - Short paper: It is not yet a completed full paper, but instead a quick report of the partial result obtained at the preliminary stage as well as the knowledge obtained from the said result. As a general rule, the authors are requested to summarize a paper within four pages.
  - System development paper: It is a paper that is a combination of existing technology or it has its own novelty in addition to the novelty and utility of an ordinary paper, and the development results are superior to conventional methods or can be applied to other systems and demonstrates new knowledge. As a general rule, the authors are requested to summarize a paper within eight pages.
  - Data Paper: A summary of data obtained in the process of a survey, product development, test, application, and so on, which are the beneficial information for readers even though its novelty is not high. As a general rule, the authors are requested to summarize a paper within eight pages.
- To submit the manuscript for ordinsry paper, short paper, system development paper, or data paper, at least one of the authors must be a member or a student member of the society.
- We prohibit the duplicate submission of a paper. If a full paper, short paper, system development paper, or data paper with the same content has been published or submitted to other open publishing forums by the same author, or at least one of the co-authors, it shall not be accepted as a rule. Open publishing forum implies internal or external books, magazines, bulletins and newsletters from government offices, schools, company organizations, etc. This regulation does not apply to a preliminary draft to be used at an annual meeting, seminar, symposium, conference, and lecture meeting of our society or other societies (including overseas societies). A paper that was once approved as a short paper and being submitted again as the full paper after completion is not regarded as a duplicate submission.

(2) Submission stage of a paper

- Delete all author information at the time of submission. However, deletion of reference information is the author's discretion.
- At first, please register your name on the paper submission page of the following URL, and then log in again and fill in the necessary information. Use the "Style Format" to upload your manuscript. An applicant should use PDF format (converted from dvi of TeX or MS Word

format) for the manuscript. As a rule, charts (figures and tables) shall be inserted into the manuscript to use the "Style Format". (a different type of data file, such as audio and video, can be uploaded at the same time for reference.)

http://www.editorialmanager.com/iieej/

- If you have any questions regarding the submission, please consult the editor at our office.

Contact:
Person in charge of editing
The Institute of Image Electronics Engineers of Japan
3-35-4-101, Arakawa, Arakawa-Ku, Tokyo 116-0002, Japan
E-mail: hensyu@iieej.org
Tel: +81-3-5615-2893, Fax: +81-3-5615-2894

2. Review of Papers and Procedures

(1) Review of a paper

- A manuscript is reviewed by professional reviewers of the relevant field. The reviewer will deem the paper "acceptance", "conditionally acceptance" or "returned". The applicant is notified of the result of the review by E-mail.
- Evaluation method

  Ordinary papers are usually evaluated on the following criteria:
  - ✓ Novelty: The contents of the paper are novel.
  - ✓ Utility: The contents are useful for academic and industrial development.
  - ✓ Reliability: The contents are considered trustworthy by the reviewer.
  - ✓ Comprehensibility: The contents of the paper are clearly described and understood by the reviewer without misunderstanding.

  Apart from the novelty and utility of an ordinary paper, a short paper can be evaluated by having a quickness on the research content and evaluated to have new knowledge with results even if that is partial or for specific use.

  System development papers are evaluated based on the following criteria, apart from the novelty and utility of an ordinary paper.
  - ✓ Novelty of system development: Even when integrated with existing technologies, the novelty of the combination, novelty of the system, novelty of knowledge obtained from the developed system, etc. are recognized as the novelty of the system.
  - ✓ Utility of system development: It is comprehensively or partially superior compared to similar systems. Demonstrates a pioneering new application concept as a system. The combination has appropriate optimality for practical use. Demonstrates performance limitations and examples of performance of the system when put to practical use.

  Apart from the novelty and utility of an ordinary paper, a data paper is considered novel if new deliverables of test, application and manufacturing, the introduction of new technology and proposals in the worksite have any priority, even though they are not necessarily original. Also, if the new deliverables are superior compared to the existing technology and are useful for academic and industrial development, they should be evaluated.

(2) Procedure after a review

- In case of acceptance, the author prepares a final manuscript (as mentioned in 3.).
- In the case of acceptance with comments by the reviewer, the author may revise the paper in consideration of the reviewer's opinion and proceed to prepare the final manuscript (as

mentioned in 3.).

- In case of conditional acceptance, the author shall modify a paper based on the reviewer's requirements by a specified date (within 60 days), and submit the modified paper for approval. The corrected parts must be colored or underlined. A reply letter must be attached that carefully explains the corrections, assertions and future issues, etc., for all of the acceptance conditions.
- In case a paper is returned, the author cannot proceed to the next step. Please look at the reasons the reviewer lists for the return. We expect an applicant to try again after reviewing the content of the paper.

(3) Review request for a revised manuscript

- If you want to submit your paper after conditional acceptance, please submit the reply letter to the comments of the reviewers, and the revised manuscript with revision history to the submission site. Please note the designated date for submission. Revised manuscripts delayed more than the designated date be treated as new applications.
- In principle, a revised manuscript willl be reviewed by the same reviewer. It is judged either accceptance or returned.
- After the judgment, please follow the same procedure as (2).

3. Submission of final manuscript for publication

(1) Submission of a final manuscript

- An author, who has received the notice of "Acceptance", will receive an email regarding the creation of the final manuscript. The author shall prepare a complete set of the final manuscript (electronic data) following the instructions given and send it to the office by the designated date.
- The final manuscript shall contain a source file (TeX edition or MS Word version) and a PDF file, eps files for all drawings (including bmp, jpg, png), an eps file for author's photograph (eps or jpg file of more than 300 dpi with length and breadth ratio 3:2, upper part of the body) for authors' introduction. Please submit these in a compressed format, such as a zip file.
- In the final manuscript, write the name of the authors, name of an organizations, introduction of authors, and if necessary, an appreciation acknowledgment. (cancel macros in the Style file)
- An author whose paper is accepted shall pay a page charge before publishing. It is the author's decision to purchase offprints. (ref. page charge and offprint price information)

(2) Galley print proof

- The author is requested to check the galley (hard copy) a couple of weeks before the paper is published in the journal. Please check the galley by the designated date (within one week). After making any corrections, scan the data and prepare a PDF file, and send it to our office by email. At that time, fill in the Offprint Purchase Slip and Copyright Form and return the scanned data to our office in PDF file form.
- In principle, the copyrights of all articles published in our journal, including electronic form, belong to our society.
- You can download the Offprint Purchase Slip and the Copyright Form from the journal on our homepage. (ref. Attachment 2: Offprint Purchase Slip, Attachment 3: Copyright Form)

(3) Publication

- After final proofreading, a paper is published in the Academic journal or English transaction (both in electronic format) and will also be posted on our homepage.

IIEEJ Trans. on Image Electronics and Visual Computing

Vol. 11, No. 1 2023

Editor in Chief: Osamu Uchida
The Institute of Image Electronics Engineers of Japan
3-35-4-101, Arakawa, Arakawa-ku, Tokyo 116-0002, Japan

Print: ISSN 2188-1898
Online: ISSN 2188-1901
CD-ROM: ISSN 2188-191x