

Studies on Stepwise Transfer of Domain
Knowledge for Computer-aided Diagnosis in
Pathology Using Deep Neural Networks

March 2019

Jia Qu

Studies on Stepwise Transfer of Domain
Knowledge for Computer-aided Diagnosis in
Pathology Using Deep Neural Networks

Graduate School of Systems and Information

Engineering

University of Tsukuba

March 2019

Jia Qu

Contents

Chapter 1. Introduction.....	1
Chapter 2. Cancer and Pathology Diagnosis	4
2.1 Cancer and Diagnostic Methods.....	4
2.2 Pathology Diagnosis.....	6
2.3 Computer-Aided Diagnosis in Pathology	8
Chapter 3. Pathology Image Recognition Using Deep Learning.....	11
3.1 Pathology Image Recognition.....	11
3.2 Deep Learning in Image Recognition	13
3.3 Pathology Image Recognition Using Deep Learning.....	14
3.4 Problems in Pathology Image Recognition Using Deep Learning.....	15
Chapter 4. Proposed Method	17
4.1 Transfer Learning	17
4.1.1 Foundation.....	17
4.1.2 Difficulties in Implementing Transfer Learning	19
4.2 Stepwise Transfer of Domain Knowledge for Computer-aided Diagnosis in Pathology.....	19
Chapter 5. Transfer of Pathological Knowledge Based on Human Observation	22
5.1 Concept and Formation.....	22
5.2 Tissue-wise “Medium-Level” Data	24
5.3 Experiments and Discussion.....	25

5.4 Summary	31
Chapter 6. Transfer of Pathological Knowledge Based on Automatic Measurement with Human Understanding	32
6.1 Concept and Formation.....	32
6.2 Automatically Generated Cell-wise “Medium-Level” Data	33
6.2.1 Processing Flow	33
6.2.2 Image Pre-processing	34
6.2.3 CILAC (Color Index Local Auto-Correlation).....	37
6.2.4 Dimensionality Reduction and Unsupervised Clustering.....	39
6.3 Experiments and Discussion.....	39
6.4 Summary	43
Chapter 7. Stepwise Transfer of Multiple Domain Knowledge	44
7.1 Motivation	44
7.2 Stepwise Transfer of Multiple Domain Knowledge	45
7.3 Extension of Relational Domain Knowledge	47
7.4 Summary	50
Chapter 8. Conclusion	52
References	55
Ethical approvals	63
List of publications.....	64
Acknowledgements.....	66

List of tables

Table 5.3.1 Datasets used in experiments (tissue-wise)	26
Table 5.3.2 Performances of the proposed two-stage knowledge transfer using tissue-wise data.	29
Table 6.3.1 Datasets used in experiments (cell-wise)	40
Table 6.3.2 Performances of the proposed two-stage knowledge transfer using cell-wise “medium-level” data.....	42
Table 7.2.1 Datasets used in experiments (multiple domain)	46
Table 7.2.2 Performances of schemes using datasets with fixed number of images.....	47
Table 7.3.1 Datasets including gastric pathology images and lymph pathology images within relational domain	48
Table 7.3.2 Performances of schemes using datasets with fixed number of images.....	50

List of figures

Figure 2.1.1 Percentage of major diseases by cause of death in Japan	5
Figure 2.1.2 Annual trend of deaths of major diseases in Japan.....	5
Figure 2.2.1 Procedure of pathology diagnosis.....	7
Figure 2.3.1 Whole slide image (WSI).....	9
Figure 2.3.2 Computer-aided diagnosis in pathology	9
Figure 3.2.1 Winning methods of object classification task in previous ImageNet Large Scale Visual Recognition Competition (ILSVRC)	13
Figure 3.2.2 Framework of image recognition using CNN	14
Figure 3.3.1 The champion approach using deep learning in CAMELYON 2016.....	15
Figure 4.1.1 Fundamental framework of transfer learning	18
Figure 4.2.1 Basic concept of the proposed stepwise transfer of domain knowledge for computer-aided diagnosis in pathology	20
Figure 5.1.1 Practical procedure of the proposed stepwise transfer of domain knowledge for computer-aided diagnosis in pathology	23
Figure 5.2.1 Stroma and epithelium appear in gastric pathology image	25
Figure 5.3.1 “Non-professionally” annotated tissue-wise datasets	26

Figure 5.3.2 Deep neural networks employed in experiments	28
Figure 5.3.3 Performances of the proposed two-stage knowledge transfer using tissue-wise data presented by ROC	30
Figure 5.3.4 A set of filtered response images outputted by the stepwise trained VGG-16 model	30
Figure 6.2.1 Procedure of generating “medium-level” dataset with color index local auto-correlation (CILAC)	34
Figure 6.2.2 Generation method of 3-level images	35
Figure 6.2.3 Gray-scale HLAC mask patterns.....	38
Figure 6.2.4 CILAC mask patterns.....	38
Figure 6.3.1 Example of automatically generated “Medium-level” data.....	41
Figure 6.3.2 Performances of the proposed two-stage knowledge transfer using cell-wise “medium-level” data.....	42
Figure 7.1.1 Procedures of Transfer of Multiple Domain Knowledge	45
Figure 7.2.1 Results of transfer of multiple domain knowledge.....	46
Figure 7.3.1 Different kinds of pathology images.....	48
Figure 7.3.2 Performance of stepwise knowledge transfer scheme using pathology images of different organs	49

Chapter 1

Introduction

Cancer is acknowledged as one of the top threats to human health. Within the diagnostic methods for a variety of cancers, advanced image diagnosis technologies such as computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET), along with more and more credible biomarkers are evolving rapidly and becoming broadly used. Meanwhile, pathology diagnosis is still realized as the gold standard to finally assess cancer's presence or absence, type, and malignance degree.

Pathology diagnosis may be conducted before or after the surgical procedure, or sometimes as an intraoperative examination as well. However, since the examination in all cases can only be performed by pathologists, the number of pathologists is showing a huge shortage to respond all of the needs. Since more than a decade ago, new technologies such as digital pathology has widely spread and facilitated faster and cheaper diagnosis due to its operational ease. Nevertheless, because the new device is still unable to alter the fact that the pathology diagnosis completely depends on the pathologist's observation and judgement, diagnosis correctness and pathologist workload alleviation remain challenges.

In order to give more substantial solution to help reduce the burden of pathologists, more and more researchers are concentrating on further assistance. Among the researches, computer-aided diagnosis (CAD), such as pre-diagnosis “screening” and pro-diagnosis “double check” based on image classification technologies are greatly expected to play a key role to facilitate more smart pathology diagnosis. In earlier researches, many of the approaches using image features or tissue’s structural feature combined with statistical classifiers have presented applicability in some degree, but often subjected to the variability of pathological images. Recently, deep learning based approaches using convolutional neural networks (CNN) has been demonstrated unprecedented power, even beyond human for many of image classification tasks. Subsequently, high expectations of computer-aided diagnosis (CAD) for pathology image has been placed on these deep learning based approaches. However, because of extremely high cost of pathologist’s professional work, the lack of well annotated pathological image data to train deep neural networks is currently becoming a new problem.

Aiming at the issue above, this dissertation introduces the concept of stepwise transfer learning, making use of a pre-trained model for the task of another relational domain is often adopted, with the goal of improving the lack of data when adopting pathological image recognition using deep learning based approaches. Even in the case using a limited number of well-annotated pathology images, the proposed stepwise transfer learning manages to establish a systematic procedure to gradually transfer the domain knowledge of general object recognition to the domain of pathology image recognition, so that construct a more specialized deep learning model and boost the recognition performance. According to our investigation, it is found that newly-introduced knowledge can be transferred in the form contained on the carrier of unannotated pathology image data. One kind is supposed to be acquired based on human observation, while another kind is considered available to obtain based on automatic measurement according to human understanding.

In accordance with the concepts above, the remainder of this dissertation is organized as follows. The relevant materials of the background of research, starting from the brief introduction of cancer and its diagnostic methods, will be presented in Chapter 2. In Chapter 3 researches of pathology image recognition using deep learning and the main challenges will be introduced. Chapter 4 will give clear explanation of the concept of the proposed scheme, involving the basis of transfer

learning and the structural design of the stepwise transfer of domain knowledge for CAD in pathology. Next, in Chapter 5 and 6, statements of how to obtain and transfer the newly-introduced pathology-relevant knowledge in a stepwise way will be unfolded and described in detail. In both of the chapters, several representative deep learning models will be employed together with other experimental materials. The effectiveness of the proposed scheme will be discussed in separated cases. Following in Chapter 7, this dissertation also explore and discuss about the feasibility of the extended form of the proposed scheme, for example, one of the most concerned form is the one when multiple domains of knowledge are used and transferred within a whole process. Chapter 8 follows at the end of the dissertation and gives the conclusions according to all the methodologies and experimental results. Moreover, based on the results present solid evidence for the effectiveness denoted in the foregoing chapters.

Chapter 2

Cancer and Pathology Diagnosis

2.1 Cancer and Diagnostic Methods

Cancer is acknowledged as one of the top threats to human health. According to the International Agency for Research on Cancer (IARC), in 2012, there were approximately 14.1 million new cancer cases and 8.2 million deaths around the world [1]. This number is estimated to increase to 24 million by 2035 and the deaths will continually rise. In Japan, according to the data provided by government in 2015 (Figure 2.1.1, [2]), cancer (or called malignant growth) is occupying 27.8% of all of the annual cancer-caused death, much more than that of heart disease (15.2%) and cerebrovascular disease (8.2%). Moreover, if we have a look at the annual trend of deaths of major diseases (Figure 2.1.2), it must be more seriously realized that compared to relatively controlled trend in the growth of deaths of other major diseases, cancer is still showing the momentum of growth. One of the reasons that best explains this phenomenon is the continuing Japan's aging process, with the consequent increase numbers of cancer patients and death.

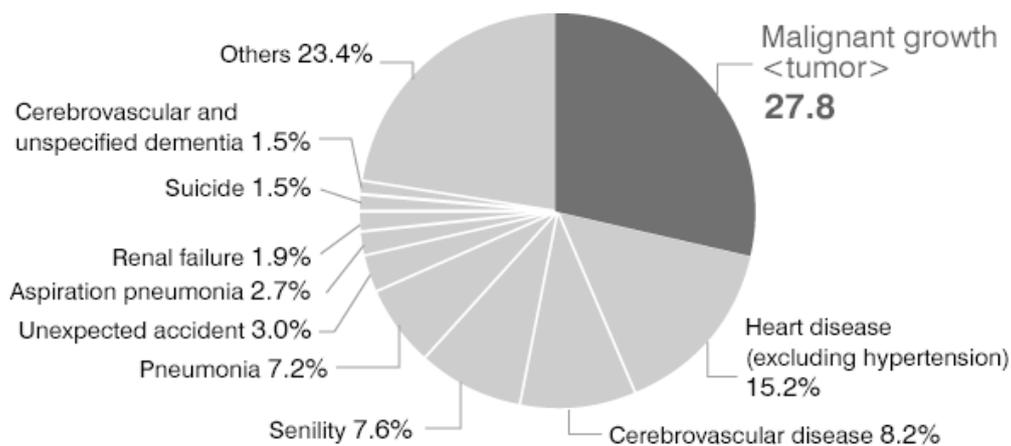


Figure 2.1.1 Percentage of major diseases by cause of death in Japan.

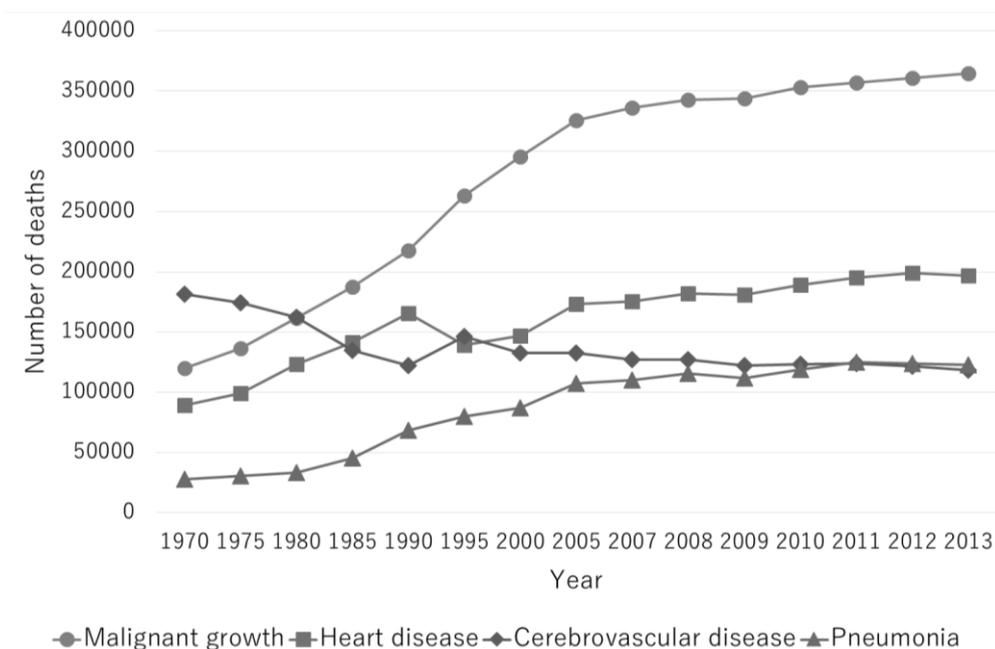


Figure 2.1.2 Annual trend of deaths of major diseases in Japan.

Aiming at diagnosing various kinds of cancers, more and more approaches have been developed. Commonly used approaches [3] can be categorized as below.

(1) *Physical exam*

Medical doctor may look for abnormalities by observing the changes in skin color or touching the enlargement of an organ. Physical exam is usually performed as the most preliminary diagnosis step.

(2) *Laboratory tests*

Biomarkers sampled from the urine and blood of the patient often help to identify the occurrence of cancer. For example, abnormal rise CA19-9

(carbohydrate antigen 19-9) may suggest pancreatic cancer, colon cancer or biliary tract tumor, while another type of biomarker CEA (Carcinoembryonic antigen) is checked when colorectal cancer, lung cancer and breast cancer are suspected. More than the biomarkers, a common blood test called complete blood count also belongs to laboratory test. Since the complete blood count may reveal an unusual number or type of white blood cells, it is deemed as an effective diagnosis approach of leukemia.

(3) Imaging tests

Well-known advanced image test technologies such as X-ray, computed tomography (CT), magnetic resonance imaging (MRI), Ultrasound and positron emission tomography (PET) makes accurate and noninvasive examination of organs possible. In many cases when cancers are in early stages before the patient has symptoms, image test technologies is used as the most powerful means to screen the small cancer that has not spread.

(4) Biopsy

Biopsy involves taking a small piece of tissue sample so that it can be examined under a microscope. Generally, two types of biopsy, cytology diagnosis and pathology diagnosis [4], are included in the field of biopsy. Cytology diagnosis is usually performed to examine the individual cells, which are coming out of the body and may infer the lesion where the cell resides, for example, lung cells appearing in sputum, cells of bladder appearing in urine, etc. As opposed to cytology diagnosis, pathology diagnosis is the study of whole tissue. Due to the sampling method that tissues may be collected by puncture, endoscopy or even from the organ under operation, pathology diagnosis usually have a wider range of applications. In most cases, pathology diagnosis is treated as the gold standard to definitively diagnose cancer, and assess cancer's type and malignance degree.

2.2 Pathology Diagnosis

The procedure of pathology diagnosis is illustrated in Figure 2.2.1 [5]. It starts from tissue sampling during operation or endoscopy examination. Specimens (or organs) from the operating room are submitted to the pathological examination department as raw. Next, specimens need to be pre-prepared. Processes include fixation, taking photos, cutting off and embedding (replace fat and moisture in specimens with paraffin). The embedded specimens are then stained by special dyes.

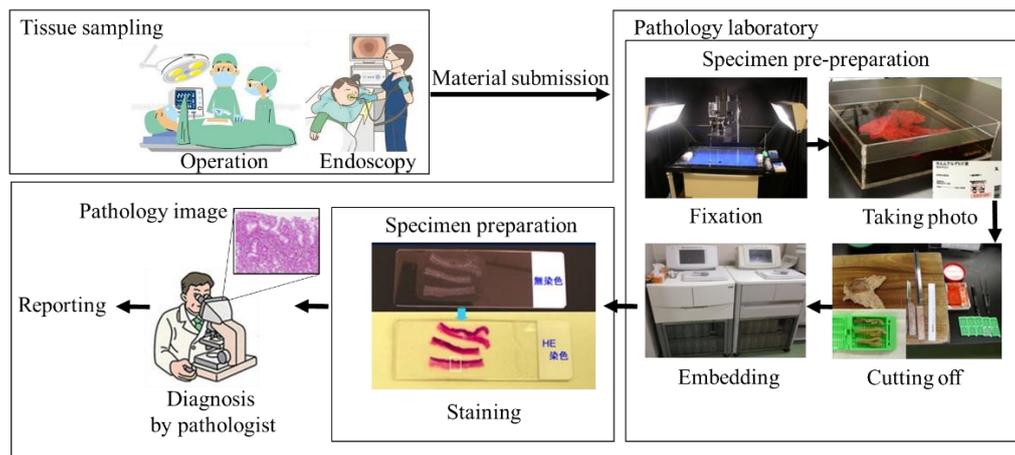


Figure 2.2.1 Procedure of pathology diagnosis.

Here, different kinds of dyes are adopted to meet different needs. The most commonly-used dye is HE (Hematoxylin-Eosin). For many of the cases, since tumors are largely classified as epithelial and mesenchymal, by making the whole tissue stained with HE (Hematoxylin-Eosin), malignant or benign tissues can be fully represented. In some other cases, cancerous cells need to be observed and accessed after the antigens are bonded with specific antibodies. These antibodies, which may be high molecular weight glycoprotein or enzyme, can be stained by corresponding dyes or have fluorescent property. Such kind of staining is named as “immunohistochemistry”. Representative staining substances used in immunohistochemistry include DAB (Diaminobenzidine, dye staining), PE (Phycoerythrin, fluorescent stain) [6], etc. After the procedures above, specimens are finally set under microscope and diagnosed by pathologists. According to pathologist’s findings, finally, reports are created and provided to the attending doctors. In this desertion, methodology centered on HE pathology images will be discussed in detail.

In earlier days, all specimens were directly observed under microscope. Although this is right suitable in the case of need for a rapid examination during a surgical procedure, there comes out a vital issue that one specimen can only be seen and understood by one pathologist once. What is more, since the specimens to be observed need to be reloaded under the microscope every time before examination, and the field of view often needs to be adjusted to fit the focus and location, it was deemed waste of diagnostic time. In a later period, as a transitional solution, digital microscope [7] was developed. It that enables specimens to be observed on a display through a digital camera mounted on microscope. In terms of configuration, the

eyepiece section, which was indispensable for the conventional microscope, has been omitted. Such kind of design has given answer to the simultaneous observation, data sharing and storage, nevertheless, the problem of slide loading is still not fully solved. Since the end of 1990s, another technology named virtual microscope or virtual slide has been developed [8]. Virtual microscope is a scanner system that can load tens of tissue specimens once and transform them into digital images at high speed and high resolution. On the PC side, the pathologist can easily perform arbitrary enlargement / reduction at any arbitrary position from macro to micro of the entire tissue sample and freely observe any part of the virtual microscope image (also known as whole slide image, WSI). Presently, virtual microscope is widely used and has shown its power in more than one aspects, such as remote rapid pathologic diagnosis, education of pathologists and diagnosis standardization.

2.3 Computer-Aided Diagnosis in Pathology

New technology have indeed brought convenience and innovation to pathological diagnosis. However, another severe problem, the shortage of pathologists, is now representing as great restriction to pathologic diagnosis and causes social problems. In the United States, the lack of pathologist workforce is become more and more concerned [9]. In japan, the number of pathologist normalized by the general population is even smaller than 1/3 of that in the United States (one pathologist per 19,000 people) [10]. The situation is even more severe in China. As reported, China has approximately one pathologist per 74,000 people [11]. On one hand, such serious shortage is now consequently leading to immense working burden on pathologists and possible errors and oversights in diagnosis. On the other hand, the number of pathological diagnosis cases continues to increase as the number of patient increases. Compared with in 2005, the number of all pathological diagnosis cases and intra-operative pathological diagnosis cases in Japan has respectively risen up to 1.7 times and 3 times by 2012. Until 2017, there were approximately 30 million cases of annual pathological examination in Japan. On contrast, there were only 2,483 pathologists approved and registered at the Japan Pathological Society [12]. From these numbers, it can be inferred that the work volume per pathologist is about 33 cases / day. Assuming that the 10 specimens need to be examined for each case, one pathologist need to observe and drawing conclusions about 330 microscopic specimens every day. Moreover, whole slide

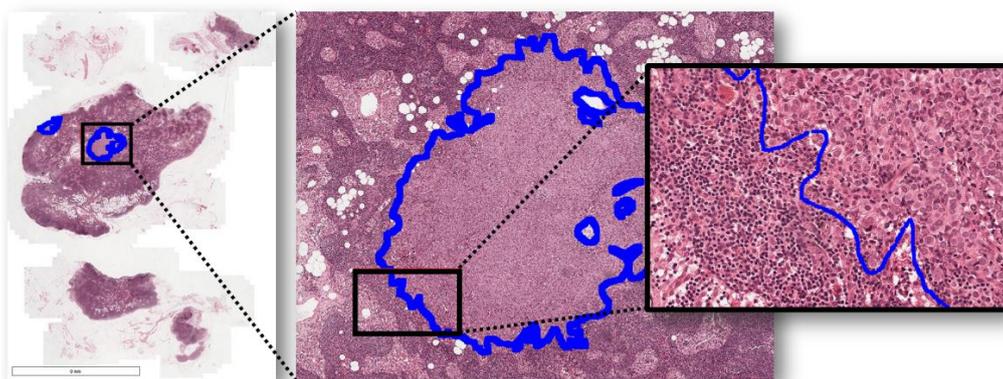


Figure 2.3.1 Whole slide image (WSI)

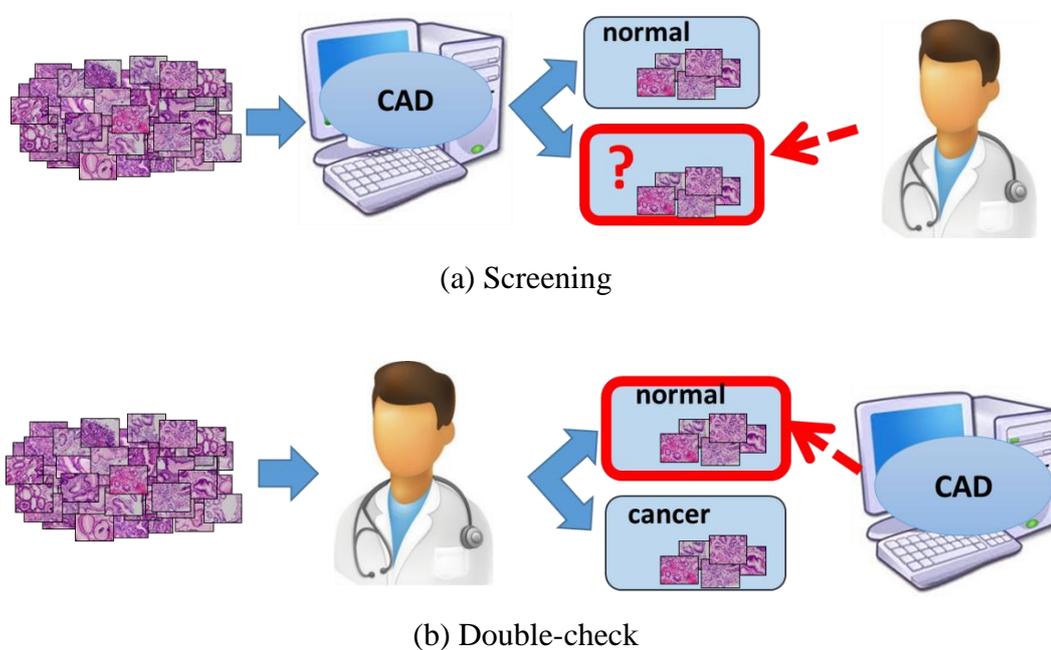


Figure 2.3.2 Computer-aided diagnosis in pathology.

image sometimes own a huge size of more than 1 billion pixels (Figure 2.3.1). Observing such images poses a challenge to pathologists, and diagnosis quality as well.

In order to reduce the burden of pathologists and improve the quality and efficiency of the diagnosis, computer-aided diagnosis is expected to play a key role to facilitate more advanced pathology diagnosis. In simple terms, computer-aided diagnosis benefits from pattern recognition technologies may automatically judge the presence or absence of cancer in the image and suggest the position if there cancer exists.

Specifically, computer-aided diagnosis can mainly help in two forms: screening and double check (Figure 2.3.2) Screening is performed prior to the doctor's diagnosis. When doctors begin to observe pathology images, it can prompt doctors to quickly prioritize the images and areas with high suspiciousness in order to get diagnostic results as quickly as possible. This is also considered to be an effective solution to shorten the pathologist's diagnosis time and reduce pathologist's workload. Double-check is used to rescreen pathology images that have been diagnosed by a pathologists, and ensure that there are no undiscovered anomalies in these images. Double-check is regarded able to help control the quality of the diagnosis, especially in high-intensity working conditions, and can also be used in the training of inexperienced pathologists.

Chapter 3

Pathology Image Recognition Using Deep Learning

3.1 Pathology Image Recognition

Many of early pathological image recognition methods employed specified histologically-concerned features or generalized texture image features. The specific histologically-concerned features, such as nuclei's area and nuclei-cytoplasmic ratio (N/C) etc. are subtly calculated from unknown images [13]. These features are compared with predefined criteria to judge whether the target image is benign or malignant. Unfortunately, such process usually meets a big issue that it is a hard task to make adequate definition for the morphological characteristics, because cancerous cells usually lack control for regular division. Thus, shape extraction failures for cells could become a direct reason for classification failures.

After 2000s, approaches using artificial image feature quantities and classifiers based on statistical models to classify a variety kinds of images became a mainstream trend. Many of the approaches have shown their effectiveness.

Compared to methods using histologically-concerned features, artificial image features seemed able to provide more robustness and adaptability to various pathological tissue appearance. For instance, one of the focused texture feature is grey-level co-occurrence matrix (GLCM). For example, Esgiar et al. [14] employed GLCM to obtain texture features corresponding to contrast, entropy, angular second moment, dissimilarity and correlation from colon biopsy, and employing linear discriminate analysis (LDA) and k-nearest neighbour algorithm (KNN) to realize the categorization of normal and cancerous colon mucosa. Likely, James Diamond et al. [15] employed Haralick features (a kind of texture features developed from GLCM) for identifying tissue abnormalities in prostate pathology images. Another mighty rival is local binary patterns (LBP). In the study of Masood et al. [16], a scheme consisting of LBP and support vector machines (SVM) are proposed and demonstrated effective for colon pathology images. In another work, O. Sertel et al. [17] developed a classification methods for neuroblastoma H&E stained whole-slide images, using co-occurrence statistics and local binary patterns similar as the above study. A recent report by Kather et al. [18] gave a relatively comprehensive investigation of texture analysis for colorectal cancer histology image. Besides of LBP and GLCM, Lower-order and higher-order histogram features, Gabor filters and Perception-like features are involved as well. In our earlier studies [19], another texture features called Higher-order Local Auto Correlation (HLAC) bonded with linear statistical models such as principal component analysis (PCA) based subspace method, were also demonstrated capable to indicate the anomaly degree of gastric pathology images. Apart from straightforward benign/malignant classification, some other methods in pathology image domain have been put forward with texture features as well, to settle similar classification-correlative tasks such as gland segmentation and grade estimation [20, 21, 22, 23, 24, 25].

While all of these texture-feature-based approaches shown promising feasibility, intractable issues still existed between the research and practical application. One particular instance is that confirming how suitable the hand-crafted geometric features are for certain tasks is quite difficult [26]. Meanwhile, the uneven H&E staining among images brings adverse impact on classification performance and makes the tasks more challenging [27, 28, 29].

3.2 Deep Learning in Image Recognition

2012 is the year when it became a breakthrough of deep learning. In the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) held in the this year, the group of Prof. Hinton of Toronto University, using the 8-layer Convolutional Neural Networks (CNN) [30] to classify 1,000 categories of objects, won over all the other teams using non-deep-learning approaches with over 10% the error rate of discrimination. After that, image recognition related researches shifted from the conventional image feature based approaches to the deep learning based approaches in a stroke. Furthermore, since 2015, newer deep learning approaches even achieved a higher recognition accuracy than the human average level (Figure 3.2.1).

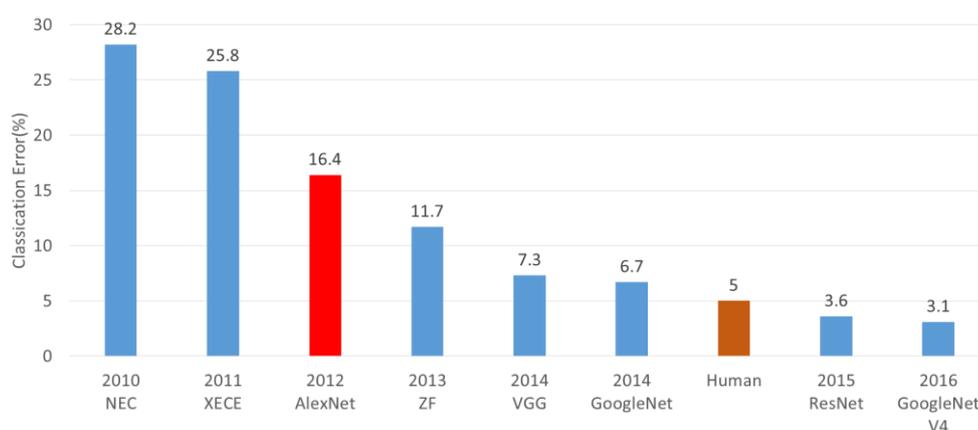


Figure 3.2.1 Winning methods of object classification task in previous ImageNet Large Scale Visual Recognition Competition (ILSVRC).

Convolutional Neural Networks (CNN) is currently the most remarkable success. The prototype of CNN can be found in Neocognitron [31] devised based on the neurophysiological findings on the visual cortex of living organism's brain. It is a neural network that alternately arranges a convolution layer corresponding to the cells for feature extraction and a pooling layer corresponding to the cells having a function to allow positional deviation hierarchically (Figure 3.2.2, [32]). Intuitively, it can be interpreted as a network that takes co-occurrence of adjacent features on different scales little by little and selectively gives information effective for identification to upper layers. Practically, refinement of such information is usually implemented by minimizing the cost function (Equation 1):

$$L(W) = \frac{1}{N} \sum_{i=1}^N l(y(x_i; W), y_i) \quad (1)$$

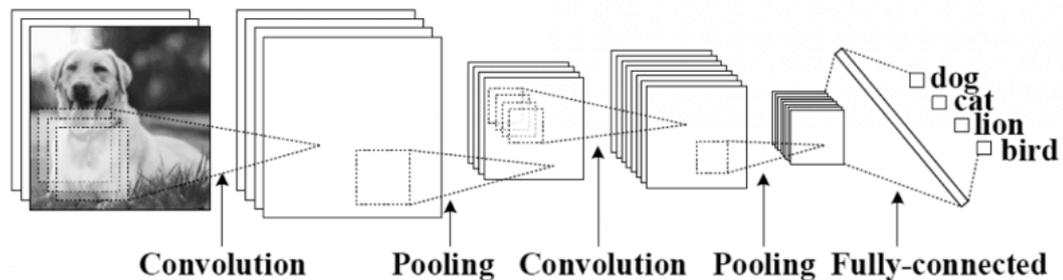


Figure 3.2.2 Framework of image recognition using CNN.

In Equation 1, while $L(W)$ indicates the total cost (difference between prediction upon current configuration and the ground truth) over a dataset of N training samples, corresponding to weights W . y_i denotes the label of training data x_i . $y(x_i; W)$ is the predicted label of x_i , while l is the lost function.

Due to more domain agnostic approach combining both feature discovery and implementation to maximally discriminate between the classes of interest [33], deep learning shows unprecedented adaptability for various kinds of images [34, 35, 36]. Accordingly, high hope are placed on deep learning to exert great power in pathology image and other medical image fields [37, 38, 39, 40, 41].

3.3 Pathology Image Recognition Using Deep Learning

Specifically within the pathology image domain, many researchers have been inspired to develop deep learning based approaches of classification and segmentation [42, 43, 44, 28, 45, 46]. Worldwide contests, such as “CAMELYON”, the goal of which is to evaluate new and existing algorithms for automated detection of metastases in hematoxylin and eosin (H&E) stained whole-slide images of lymph node sections [47], have attracted the participation of many top researchers. In 2016, around 70% of the participating methods took advantage of deep learning based approaches, leaving a small number of those using classical machine learning approaches. One of the representative approach, the 1st place one submitted by Harvard Medical School and Massachusetts Institute of Technology [48] presented a typical processing flow of pathology image recognition using deep learning. As illustrate in Figure 3.3.1. In the training phase, training data include a number of

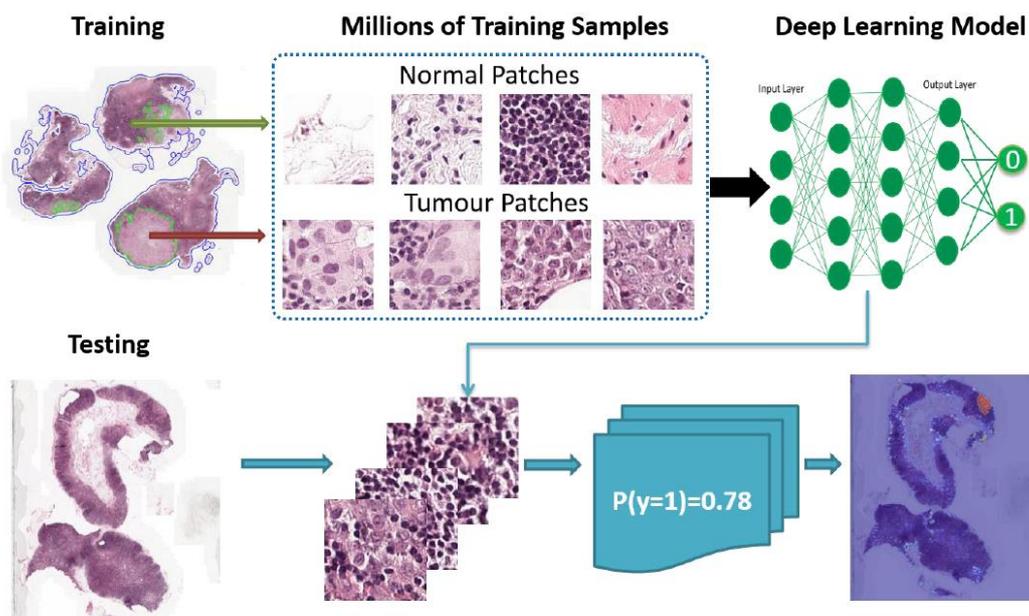


Figure 3.3.1 The champion approach using deep learning in CAMELYON 2016.

whole slide images, within which the tumor existing areas are circled to distinguish from normal areas. Both areas are divided into millions of normal patches and tumor patches. In the next, deep learning model then gets trained with these patches. Here, among several state-of-the-art deep learning models, the architecture based on GoogLeNet (developed by Szegedy et al. and won in ILSVRC 2014, [49]) has been validated as the most capable one. In the test phase, the trained deep learning model is adopted to categorize all patches of an unlearned image. According to the map of normal-tumor probability, suspected tumor area can be eventually figured out.

Apart from approaches of classification and segmentation, deep learning are also utilized in new patulous applications for pathology diagnosis. Examples include staining normalization [50], assessment of tumor proliferation [51], and comprehensive multimodal mapping between medical images and diagnostic reports [38].

3.4 Issues and Challenges

In order to adopt deep learning based approaches, large datasets are always indispensable to train more capable deep neural models and raise their performance

[52]. However, unlike natural image datasets which can be acquired based on internet and automated categorizing techniques, building up high quality pathology image datasets, anyhow, requires professional observation and annotation by pathologists. As foregoing statement, because of the lack of the number of pathologists, a tremendous work burden is placed on daily diagnosis and educational activities. Accordingly, well-annotated data usually cost vast financial resources and manpower. In this situation, how to maximize the recognition performance with limited data becomes a very realistic and urgent problem when the amount of data cannot fully meet the requirements of training deep learning networks.

Chapter 4

Proposed Method

4.1 Transfer Learning

4.1.1 Foundation

Transfer learning is a common countermeasure against training data shortage. When there is a model to be trained for the task of a certain domain, transfer learning making use of a pre-trained model for the task of another relational domain is often adopted (as shown in Figure 4.1.1). This is based on the acknowledgment that knowledge gained from the source domain is conducive to help solve the target task. Generally speaking, several advantages of transfer learning can be cited.

(1) It is possible to achieve higher recognition performance even when there are few learning data in the target domain.

(2) Compared to learning from a random initial state (learn-from-scratch), recognition performance when using transfer learning is advantageously dominant in many cases.

(3) Computational resources are saved because the learning time until equivalent

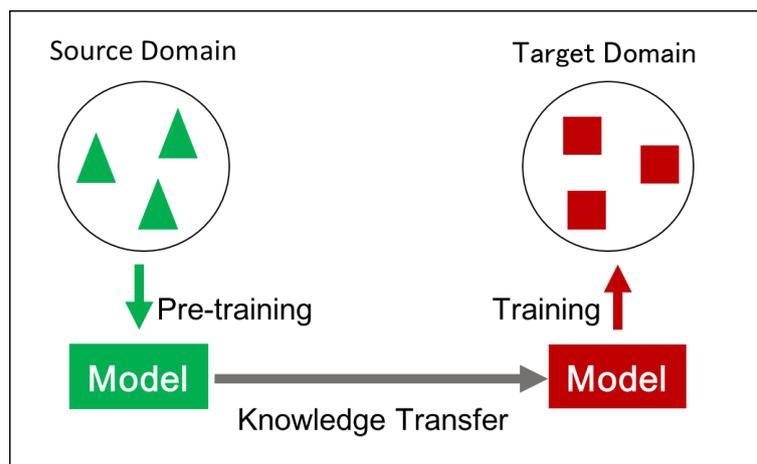


Figure 4.1.1 Fundamental framework of transfer learning.

recognition performance is short.

These advantages are due to an important trait of deep neural networks. Deep neural networks (for instance, CNN) strongly depend on its initial status, thus it is significant to obtain appropriate initialization as much as possible in order to avoid over-fitted learning or local minimum traps. Generally, the forefront layers of a CNN are considered analogous to the conventional texture features and applicable to many of related tasks, while the later layers capture more abstract image content by combining low-layer features involving more specific information corresponding to the target task [53]. Based on this fact, compared with the random initialized state of a deep learning network, the final state of part of the deep learning network pre-trained in the task of the source domain is considered relatively closed to the optimal solution for the task of the target domain, and therefore, a positive effects reflect in saving of required data amount of target domain, shortening of training time and improvement of accuracy.

Practically, it is a common way to adopt transfer learning that for a task with only a small amount of target domain data that can be used for training, using a model trained with a task with a large amount of source domain data. For instance, rather than training from scratch, researchers choose to train a deep neural network which has been pre-trained with large-scale image datasets (e.g. ImageNet) to obtain a more specialized network corresponding to target tasks can usually yield more advantageous results [54, 55, 56]. Transfer learning is also deemed particularly useful when training deep learning networks for medical image recognition.

4.1.2 Difficulties in Implementing Transfer Learning

In general image recognition tasks, training is performed on the premise that the training data and test data have the same or similar distribution, but in actual world data it is not so often. For various kinds of images, it is necessary to consider various differences of the accessible features in line with their contents. For example, the color of beak is an important characteristic when categorize a bird image, while texture and direction information is distinctive to recognize a remote sensing image. Therefore, it is a practical issue how to train properly to match the distribution of source and target on the feature space.

Actually, it is quite hard for us to understand the correlation between these tasks. This issue is exactly arising in pathology image classification. On one hand, in light of common human's perception, pathology images usually have more complicated appearances than natural images because it is difficult to figure out the intuitionistic difference between benign and malignant images at a glance due to their color uniformity of H&E stain and componential similarity of tissues. On the other hand, pathologists are able to distinguish various pathological components and structures within the image, owing to professional knowledge. Based on this knowledge, pathologists can precisely tell where abnormality is occurring. Nevertheless, natural image datasets for pre-training rarely contain relevant information. From this perspective, it is believed that it's crucial to build a bridge which can reasonably transfer the neural networks from the task of pre-training classification to the final benign/malignant judgment of the well-annotated pathological images.

4.2 Stepwise Transfer of Domain Knowledge for Computer-aided Diagnosis in Pathology

Aiming at the issue of domain adaption mentioned above, in this dissertation, stepwise transfer of domain knowledge for computer-aided diagnosis in pathology has been proposed. Similar approach adopting stepwise transfer has been presented in other domain [57]. The core motivation is that it is believed to be necessary and possible to make deep neural networks learn to understand pathology images in a rational way. With regard to the perception manner and learning progress of recognizer (which can be imagined as a human), before drawing

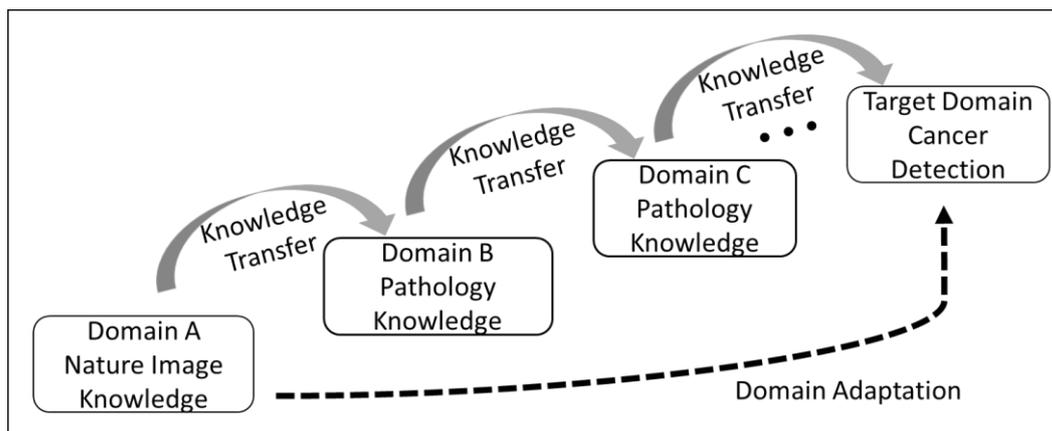


Figure 4.2.1 Basic concept of the proposed stepwise transfer of domain knowledge for computer-aided diagnosis in pathology.

conclusions of benign or malignant for a pathology image, usually, it should understand the difference of general visual characteristics, such as color and texture. Such kind of knowledges are deemed acquirable from natural image domain. Afterwards, the recognizer may concentrate more on structural and morphological characteristics, including but not limited to the spreading status and density of the specified contents which need to be understood. As to the pathology field, the size of cell, the degree of nucleus distortion, nucleus size and nuclear-cytoplasmic ratio are considered to be involved. At this step, obviously, knowledge is needed to be acquired by make the recognizer cope with a task in a more professional (or say pathology-related) domain. On account of all of the above steps, the recognizer finally become competent to classify benign and malignant images. In line with this procedure, the deep neural network is expected to gradually acquire and get optimized with the knowledge corresponding to different levels of target-task correlation, so that finally become highly adaptive to the task of target domain.

For deep neural networks themselves, this approach above is deemed able to reduce the difficulty of optimization of internal weights when re-training deep neural networks after knowledge transfer. Compared to the coping with two tasks with huge gaps, proper fine-tuning often leads to good adaptability when deep neural networks are placed in similar but non-identical tasks. This is also the premise and an important reason that the above-mentioned stepwise transfer of knowledge is considered to be applicable to deep neural networks.

Practically, as shown in Figure 4.2.1, in the case when using a limited number of well-annotated pathology images, the proposed stepwise transfer learning [58] can

manage to establish a systematic procedure starting by transferring the domain knowledge of natural image recognition. The large-scale object recognition (e.g. ImageNet) tasks is a reasonable choice. Next, according to our investigation, it is found that two types of pathology-relevant knowledge can be transferred in the form contained on the carrier of unannotated pathology image data. One kind is supposed to be acquired based on human observation, while another kind is considered available to obtain based on automatic measurement according to human understanding. After all, the well-annotated pathology images are finally employed to conduct the last optimization and finish the domain adaption procedure.

In the following parts, the proposed stepwise transfer of domain knowledge for computer-aided diagnosis in pathology will be discussed in detail. In Chapter 5, knowledge acquired based on human observation will be realized by introduce tissue-wise datasets. In Chapter 6, knowledge acquired based on automatic measurement according to human understanding will be introduced via cell-wise datasets. Moreover, in Chapter 7, I will give additional discussion on the hyperparameter of the proposed stepwise transfer scheme, including the number of steps involved, and the availability of different kinds of pathology images.

Chapter 5

Transfer of Pathological Knowledge Based on Human Observation

5.1 Concept and Formation

As mentioned in last chapter, in order to make the deep neural networks gradually acquire and transfer the knowledge corresponding to different levels of target-task correlation, and finally become adaptive to the task of target domain, naturally, deep neural networks demand to be assigned different tasks of relational domains in a proper flow. To be intuitive, a practical model of stepwise transfer of domain knowledge is shown in Figure 5.1.1. In general, the proposed stepwise transfer of domain knowledge conclude three main steps.

Before all, in order to get the common basics of image recognition, classification of large-scale ImageNet data is set as the initial task. For convenience, the knowledge to be transferred is named as “low-level” knowledge, which means its target-task correlation is relatively low. Meanwhile, the ultimate task, classification of benign/malignant pathology images, is undoubtedly placed at the last. At this

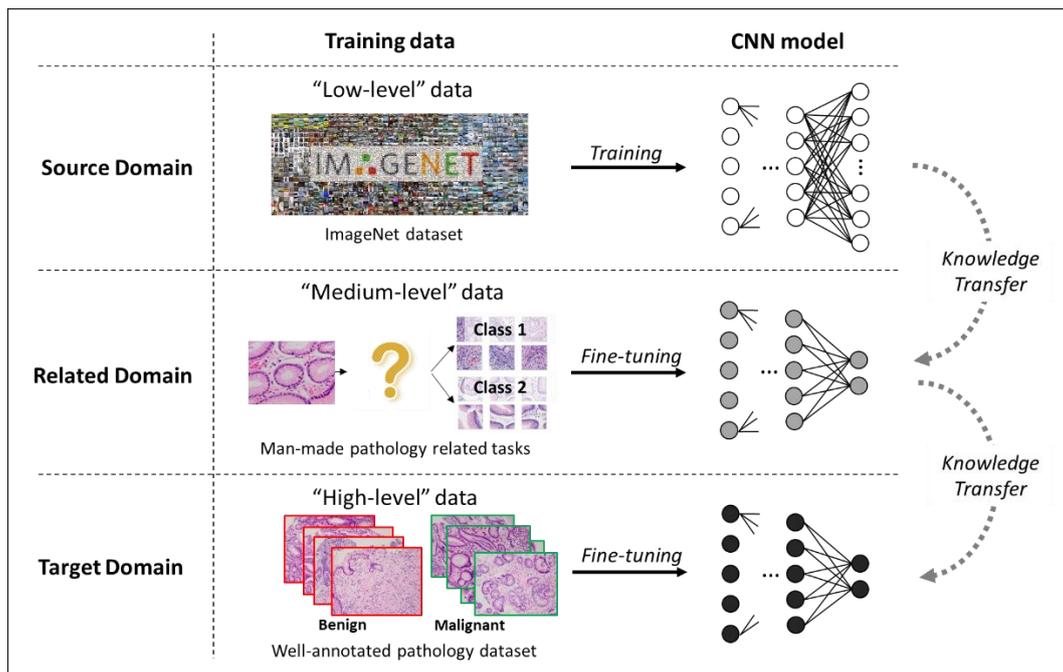


Figure 5.1.1 Practical procedure of the proposed stepwise transfer of domain knowledge for computer-aided diagnosis in pathology.

step, all of the knowledge which have been transferred from foregoing steps are combined together and directly applied to the optimization of the final task. Here, the knowledge involved is called "low-level" knowledge. Between these two domains, according to the proposed conception, within the relational domain, knowledge owing medium level target-task correlation are supposed to be learned by deep neural networks via appropriate tasks so that bridge the source domain and the target domain. In this scheme, since knowledge to be acquired and transferred are all introduced by corresponding data without exception, thus, data employed to train the deep neural networks can be called "low-level" data, "medium-data" and "high-level" data, respectively.

In terms of the configuration above, what needs to be done is to train the deep neural networks with datasets of different levels in succession. As illustrated in Figure 5.1.1, after each time knowledge transfer is done, nodes colored darker in CNN models denote that CNN has learned a more specified (deeper) representation for the pathology image classification task. When the number of output classes changes, definitely, the network architecture needs to be adjusted accordingly.

Under this framework, what needs to be solved next is to propose a task in the relevant domain to introduce relevant knowledge. As stated in Chapter 4, structural

and morphological characteristics, such as distribution of tissue, size of cell, degree of nucleus distortion, and nuclear-cytoplasmic ratio are regarded as promising clues. Therefore, modeling the task of benign/malignant classification, this dissertation will focus on characteristics that are relevant to these aspects and that are easy to be introduced via new datasets and tasks.

As for the “medium-level” datasets, it is necessary to meeting the prerequisite requirement that they could be acquired at much lower cost than the datasets annotated by pathologists, and meanwhile. Hence, this dissertation proposes the following two ideas. The first idea is to make use of the existing characteristics of the pathological images directly by employing the different components of the tissue structure as the natural classes. The classes utilized to create artificial tasks and achieve knowledge transfer is then totally based on human observation. The second idea is to take over image recognition tools to classify pathology images which containing different but more abstract characteristics. Afterwards, above-formed classes are chosen within the scope of human understanding to form the data to be transferred.

5.2 Tissue-wise “Medium-Level” Data

In this chapter, I would like to discuss about the specific implementation of the first above-stated idea: to make use of the existing characteristics of the pathological images directly by employing the different components of the tissue structure as the natural classes, with which artificial tasks and achieve knowledge transfer are then realized based on human observation. Based on the investigation, this paper presents a feasible stroma-epithelium datasets which can be made by non-professional annotator with only a little pathologist’s direction. Epithelium and stroma are two tissue types that can be found in every organ [59]. In gastric pathology domain, epithelial tissues line the outer surfaces of gastric mucosa, while stroma tissue locates right under epithelium. Since cancer metastasis between epithelium and stroma is deemed as inextricable to cancer’s progression [60], both of the two types of tissue are usually extracted during biopsy examinations and revealed in the pathology images. Although it is quite difficult to identify if the two types of tissue are in order for non-professional workers, actually, it is found that stating the visual difference between them is quite an undemanding work. As shown in Figure 5.2.1, the upper area encircled by dashed contour indicates epithelium

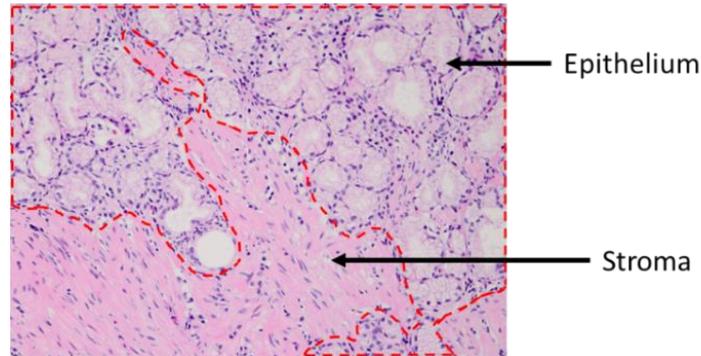


Figure 5.2.1 Stroma and epithelium appear in gastric pathology image.

tissue, while the remaining lower area denotes stroma tissue. Because of their respective functions, obviously, epithelium tissue appears as organized arrangement but stroma tissue seems scattered and disordered. Such distinct difference is considered quite beneficial to our “non-professional” work. Without pathologist’s expensive annotation, it would be encouraging if these epithelium and stroma images can impart the pathology knowledge to the deep neural networks on the basis of our assumption. Practically, I have managed to collect a large amount of images of epithelium and stroma, segment them with stroma area and epithelium and manually and separate into three classes (epithelium, stroma and background).

5.3 Experiments and Discussion

(1) Experimental Procedures

In order to evaluate the effectiveness of our proposed scheme, with the tissue-wise “medium-level” data, stepwise knowledge transfer will be implemented on several representative CNN architectures. The performances will be compared to the learn-from-scratch scheme that using well-annotated pathology data only, and the common used one-stage scheme within which knowledge transfer is directly conducted from “low-level” domain to “high-level” domain. Furthermore, this dissertation will talk over how the proposed stepwise knowledge transfer scheme performs upon well-annotated datasets of different sizes, so that to produce more evidence to validate the adaptability of the conception.

(2) Datasets

This dissertation employs three types of data including “low-level”, “medium-

Table 5.3.1 Datasets used in experiments (tissue-wise).

Data Type	Category	Training	Validation	Test
Tissue-wise Data	Background	15,000	1,000	-
	Epithelium	15,000	1,000	-
	Stroma	15,000	1,000	-
Well- annotated Data	Small	540 + 540	1620 + 1620	2,700 + 2,700
	Large	5400 + 5400		

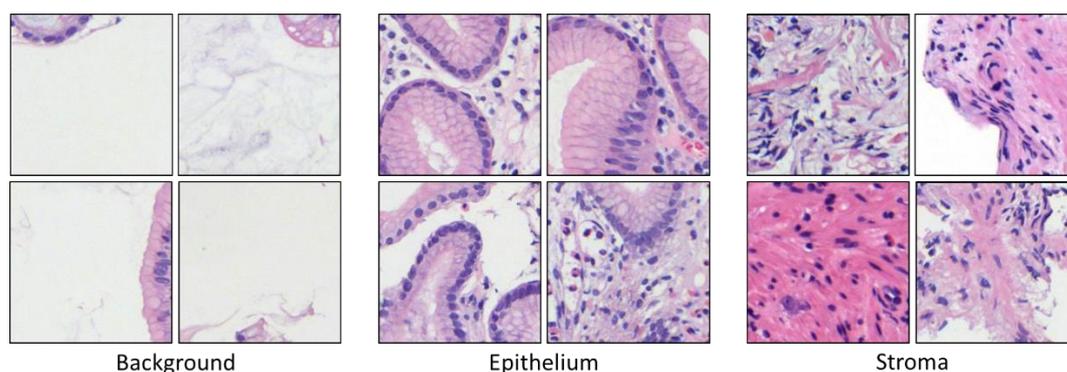


Figure 5.3.1 “Non-professionally” annotated tissue-wise datasets.

level” and “high-level” data, respectively used for the initial training, the first stage of knowledge transfer and the 2nd stage of knowledge transfer. In practice, ImageNet data [61] containing approximately 1.2 million images in 1,000 separate categories are customary utilized to initialize the CNN models. As to gastric pathology images, all the datasets utilized are illustrated in Table 5.3.1. By depicting maps of epithelium and stroma, this dissertation succeeded in collecting 48,000 tissue-wise patches (256×256) separated into “background”, “epithelium” and “stroma” categories (Figure 5.3.1). In each category, 15,000 patches are used as training data, while the remained 1,000 patches are used for validation. As to the well-annotated “high-level” datasets, in order to evaluate the efficacy and generalization of the proposed two-stage scheme, I have prepared well-annotated datasets in two different sizes. One is a small train dataset including 540 benign and

540 malignant patches. Another one is a non-augmented large dataset of 5,400 benign and 5,400 malignant patches within which the small dataset is included. Except from the former datasets, this paper additionally use a validation dataset including 1,620 benign and 1,620 malignant patches to select the best model configuration, and a test dataset of 2,700 benign and 2,700 malignant patches to finally evaluate the performance in each optional case. It is noteworthy that there is no overlap between the “medium-level” datasets and the “high-level” datasets and no overlap among the training, validation and test datasets.

(3) Deep Neural Networks Adopted in Experiments

This dissertation employs three representative deep neural networks: AlexNet [30], VGG-16 [36] and GoogLeNet (Inception V3) [49].

AlexNet, named after its developer, is the regarded as the representative CNN model. It won the ILSVRC in 2012 and led the boom of deep learning approaches. AlexNet consists of a convolution neural network with 5 convolutional layers and 3 full-connected layers (Figure 5.3.2 (a)). Currently, opportunities to use AlexNet itself have decreased, but the original concept and techniques, such as ReLU activation function, training with multi GPUs, data augmentation and dropout are still widely adopted in the subsequent deep neural networks.

VGG-16 is another famous CNN model. The 16-layer structure model (Figure 5.3.2 (b)) has high applicability to image recognition, with the features such as structural clarity and small scale. VGG-16 won the second place of image classification in ILSVRC 2014.

GoogLeNet (Inception V3) was the winner of ILSVRC 2014. The significant feature of GoogLeNet is that it defines a small micro networks called an “inception module”, which composes of multiple convolution layers and pooling layers. These inception modules build up one large CNN by overlapping it like a normal convolution layer. In the inception module, the micro networks are branched, after convolution with different sizes is performed. At the end of the module the outputs are all connected. This configuration is considered contributive to make the weight of the convolution layer sparse and improve the efficiency of optimization.

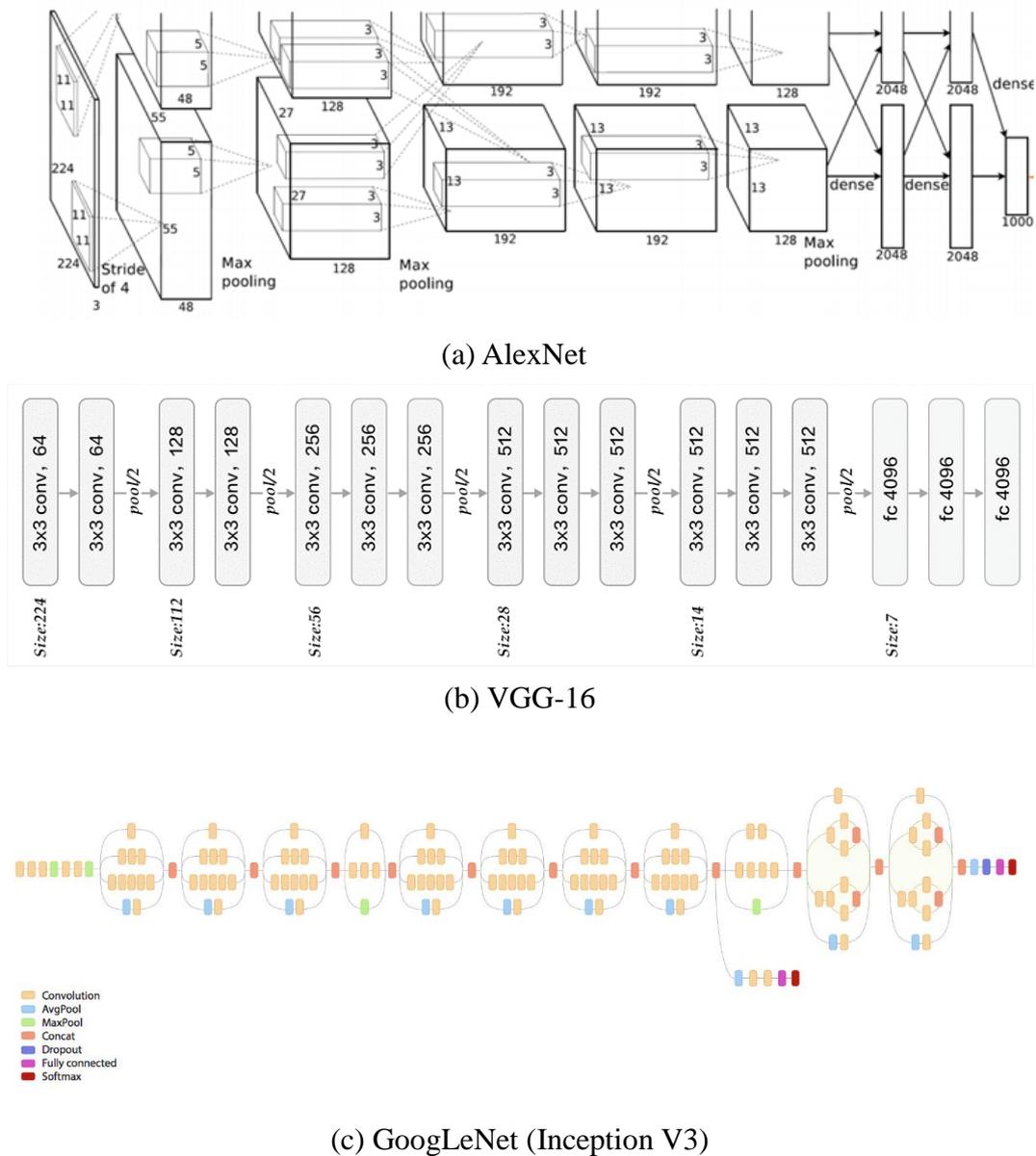


Figure 5.3.2 Deep neural networks employed in experiments.

(4) Experimental Results

This part will specifically discuss about our proposed stepwise scheme when tissue-wise data are employed for the 1st-stage knowledge transfer and well-annotated datasets are used for the 2nd-stage knowledge transfer. The performances are collected from the experiments performed with well-annotated pathology image datasets in different sizes, and different deep neural network architectures. In this paper, this paper concurrently adopts AUC, ACC, Precision and Recall as the evaluation criteria [62].

Table 5.3.2 Performances of the proposed two-stage knowledge transfer using tissue-wise data.

Data Size	Scheme	CNN Architecture											
		VGG-16				AlexNet				GoogLeNet (Inception V3)			
		AUC	ACC	Precision	Recall	AUC	ACC	Precision	Recall	AUC	ACC	Precision	Recall
Small	One stage	0.879	0.793	0.863	0.695	0.828	0.723	0.74	0.72	0.838	0.753	0.75	0.75
	Two stage(tissue)	0.914	0.829	0.865	0.781	0.844	0.761	0.76	0.76	0.877	0.772	0.79	0.77
Large	One stage	0.936	0.836	0.957	0.703	0.867	0.794	0.80	0.79	0.881	0.779	0.79	0.78
	Two stage(tissue)	0.963	0.881	0.869	0.898	0.920	0.837	0.84	0.84	0.934	0.862	0.86	0.86

Firstly, the $AUC = 0.922$ is achieved by adopting VGG-16 and learn-from-scratch approach upon the large well-annotated pathology data. In comparison, both schemes using one-stage and two-stage knowledge transfer have achieved better performance. As a reference, this result also confirms that rather than learn-from-scratch, using large-scale data to initialize deep neural networks often lead to more expected accuracy. On the basis of confirming the need to use knowledge transfer, results are discussed in correspondence to various case. In Table 5.3.2, among all of the couples of rival schemes, the proposed two-stage knowledge transfer using tissue-wise data has yield notable promotion. Specifically, in the results of small data group, AUC value is raised by 0.035, 0.016 and 0.039, when CNN architectures VGG-16, AlexNet and Inception V3, are adopted, respectively. In the large data group, the corresponding AUC values is raised by 0.027, 0.053 and 0.053. Although the performance using smaller training data are expected to be more boosted, according to AUC values, it is found that our proposed scheme has actually brought slightly more benefit to the large well-annotated dataset groups. Meanwhile, if the focus is moved to ACC values, the fact is acknowledged that the greatest improvement happens when our proposed scheme using Inception V3 is adopted upon the large dataset. The accuracy has remarkably increased from 0.779 to 0.862. Besides, precision and recall, which are commonly used for medical image classification, are appearing with the similar trend to AUC and ACC. As more intuitively illustrated in Figure 5.3.3, three CNN architectures combined with two datasets have produced six ROC Figures. The red curve denotes the two-stage scheme using tissue-wise “medium-level” dataset, while the green curve denotes the conventional one-stage scheme. It is clear at a glance, in each figure, our proposed scheme possesses overwhelming area all along both the false positive rate

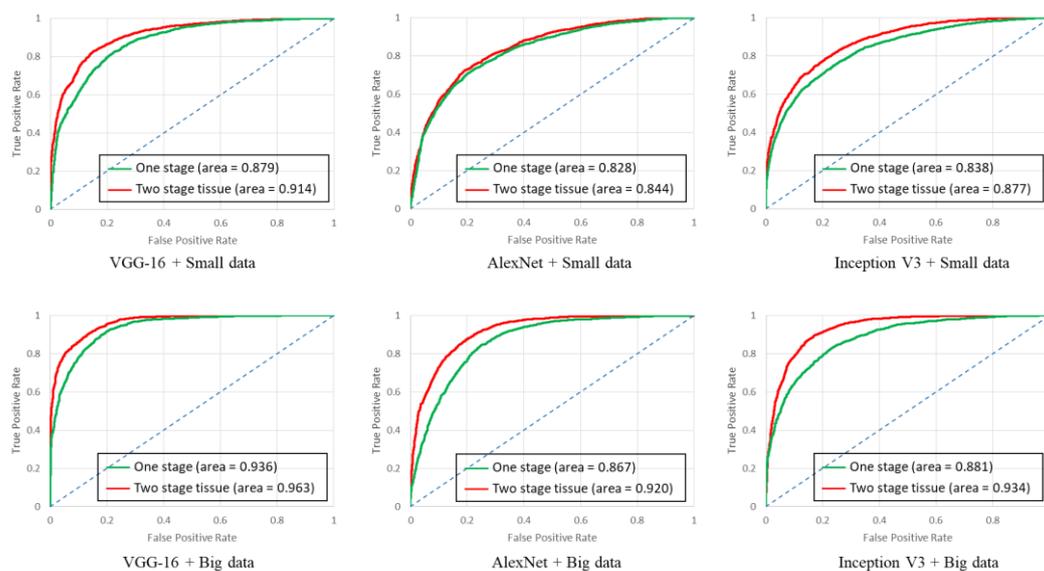


Figure 5.3.3 Performances of the proposed two-stage knowledge transfer using tissue-wise data presented by ROC.

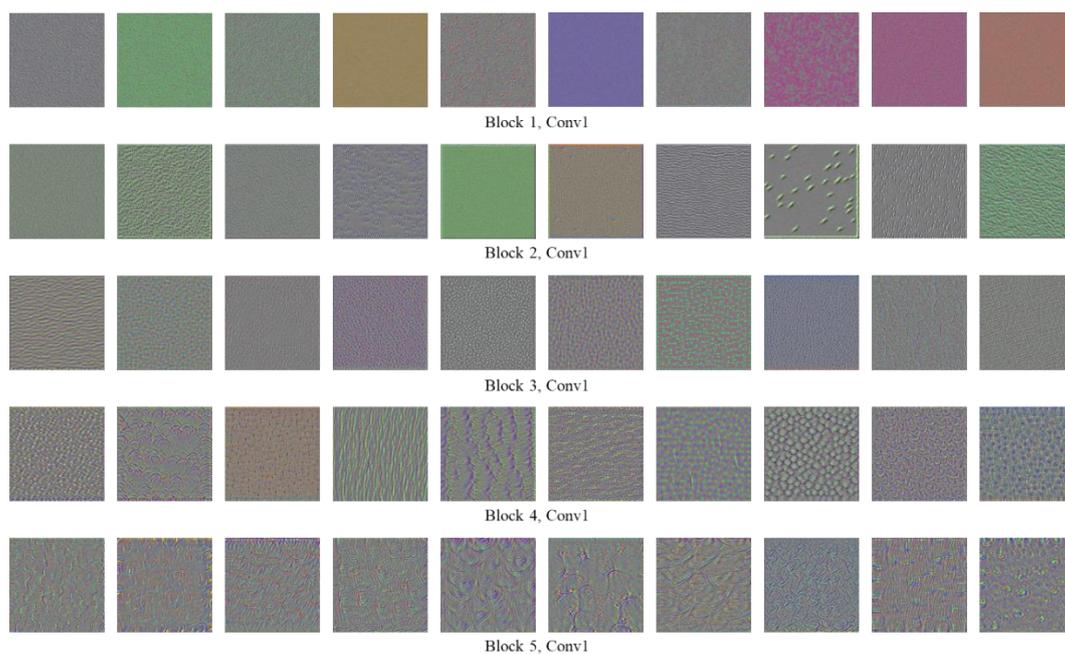


Figure 5.3.4 A set of filtered response images outputted by the stepwise trained VGG-16 model.

axis and true positive rate axis. These results have proved that our proposed scheme is adaptable and rarely dependent on the amount of well-annotated data. In addition, a set of filtered response images exported by the stepwise trained network are displayed in Figure 5.3.4 Considering the practical and intuitive facility, I

investigate on VGG-16 model and obtain some of the outcomes of the first convolutional layers from all of the convolutional blocks as shown.

In line with the common sense, larger training data yield larger absolute AUC value. Nevertheless, if this paper makes comparison between the data size crosswise, it can be noticed that when use the proposed two-stage scheme with the small dataset, the proposed stepwise scheme has actually boosted the performance up to the level approaching to that when only one-stage knowledge transfer is implemented with the large dataset which is 10 times that of the small one. By viewing the two rows “Small-Two stage (tissue)” and “Large-One stage” (right the two rows in the middle of Table 5.3.2), it is not hard to see that the differences of AUC values between the two rows are no larger than 0.023 (AlexNet), while the largest difference of ACC values is only 0.033 (AlexNet). These results can fully prove that the introduction of tissue-wise information has indeed contributed to making the deep neural networks “understand” pathology images in line with the ascending target-domain relevance. Hence, it is reasoned to infer that the proposed two-stage knowledge transfer scheme can be used as an alternative method to boost the performance when the number of well-annotated pathology image data is limited, but “non-professional” annotation is practicable even simply based on human observation.

5.4 Summary

This chapter has discussed about feasible type of knowledge that can be transferred to bridge the source domain and target domain. A novel idea is to make use of the existing characteristics of different components of pathological images directly based on human observation. Specifically, background-stroma-epithelium datasets is presented that can be generated by non-professional work with only a little pathologist’s direction. The datasets are utilized to create artificial tasks and conduct knowledge transfer. Comparative experiments implemented among different deep neural networks and well-annotated pathology images of two sizes, it is noticeably found that the proposed two-stage knowledge transfer scheme is most contributive to boosting the classification accuracy.

Chapter 6

Transfer of Pathological Knowledge Based on Automatic Measurement with Human Understanding

6.1 Concept and Formation

In last chapter, the first type of knowledge to be transferred using characteristics of different components of pathological images directly based on human observation has been discussed. Therefore, this chapter will concentrate on the realization of another idea: to take over image recognition tools to classify pathology images which containing different but more objective characteristics and form the data to be transferred within the scope of human understanding.

Within the tissue-wise domain, human observation and judgment objects are placed on some intuitive and easy to judge features. Such kinds of features may include presence or absence of the gland, arrangement and distribution of the cells. Meanwhile, non-professional workers only need to roughly point out where belongs

to which class. However, compared to tissue-wise information, morphological characteristics of the cells, which may include the shape and size of cell, and nuclei-cytoplasmic ratio (N/C), usually appears to reveal in a more abstract state. This is because that one can hardly tell which parts of the images should be categorized into which class without any absolute criterion. Therefore, in order to make use of these information, the proposed scheme has to introduce specialized tools for traversing a large number of images to give uniform and considerable criteria of measurement.

Another motivation for achieving the above-stated idea is that when there is a serious shortage of manpower, if the data involving knowledge to be transferred can be generated in an automatically generated method, then as the amount of unlabeled pathology increases, it is possible to generate more and more data for knowledge transfer at zero cost. At the same, if it can be proved that the involved knowledge is beneficial to improve the effect of knowledge transfer between source and target domains, and ultimately improve the accuracy of benign/malignant classification, then it will be more evadible to suppose that the scheme presented will have a broader and more determined applicability.

6.2 Automatically Generated Cell-wise “Medium-Level” Data

6.2.1 Processing Flow

In the light of our aforesaid conception, it is considered to adopt a reliable way to provide with rough and robust, but weakly pathology-related information based on automatic statistical measurement to fulfil this demand. According to my earlier study [63], color index local auto-correlation (CILAC, [64]) has been evidenced as an independently competent hand-crafted feature in pathology image classification. Notice that feature extraction with CILAC can only be implemented to indexed pathology images possess three color levels: background, nuclei and cytoplasm. These three components are deemed to contain most of the crucial information for morphological analysis. Meanwhile, because the color indexing process is equivalent to normalizing the color space in an extremely rough way, the color indexed images are regarded more robust to uneven staining intensity. Therefore,

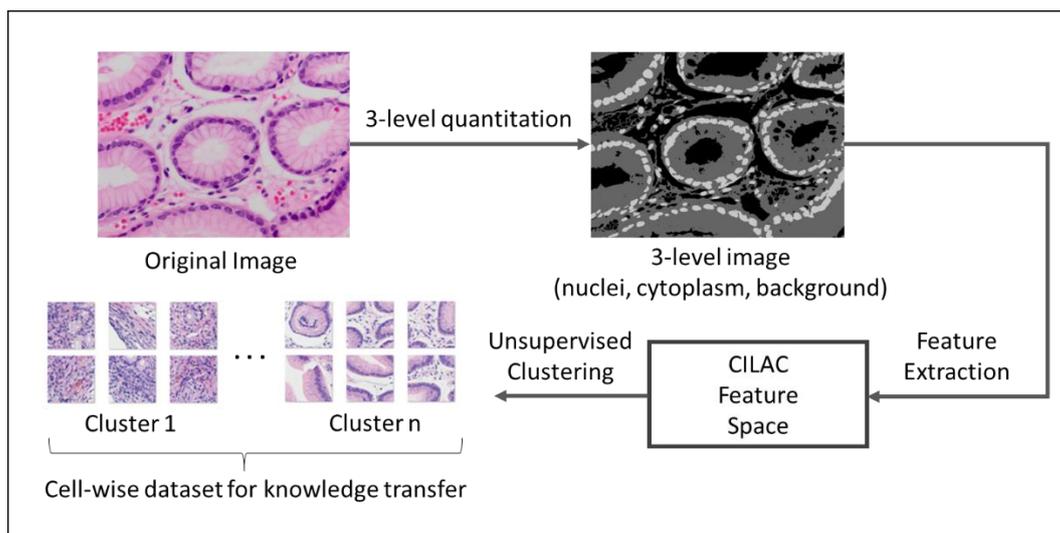


Figure 6.2.1 Procedure of generating “medium-level” dataset with color index local auto-correlation (CILAC).

in this paper, the proposed scheme takes advantage of CILAC based feature extraction on color-indexed images and expect to collect sufficient anatomical pathology information with less noise.

Accordingly, the proposed approach denoted in Figure 6.2.1, applies a string of steps including image processing, feature extraction and unsupervised clustering. First, 3-level quantitation is implemented on original pathology images so that obtain 3-level images including only three components: nuclei, cytoplasm and background. The details will be instructed in the next section. Afterwards, with the 3-level images, CILAC features are extracted to establish a feature space. Next, unsupervised clustering approach is employed to separate images into several clusters. After all, based on human observation, clusters far apart are finally selected as the cell-wise dataset for knowledge transfer.

6.2.2 Image Pre-processing

The proposed processing of pathological image avoids the influence of uneven staining to the anomaly detection. In order to concentrate on the area with high target-domain correlation, pathological images are segmented into three regions, including nuclear, cytoplasm and background. The process of this part will be explained in the following.

From the preliminary experiments, it is found that there is significant difference

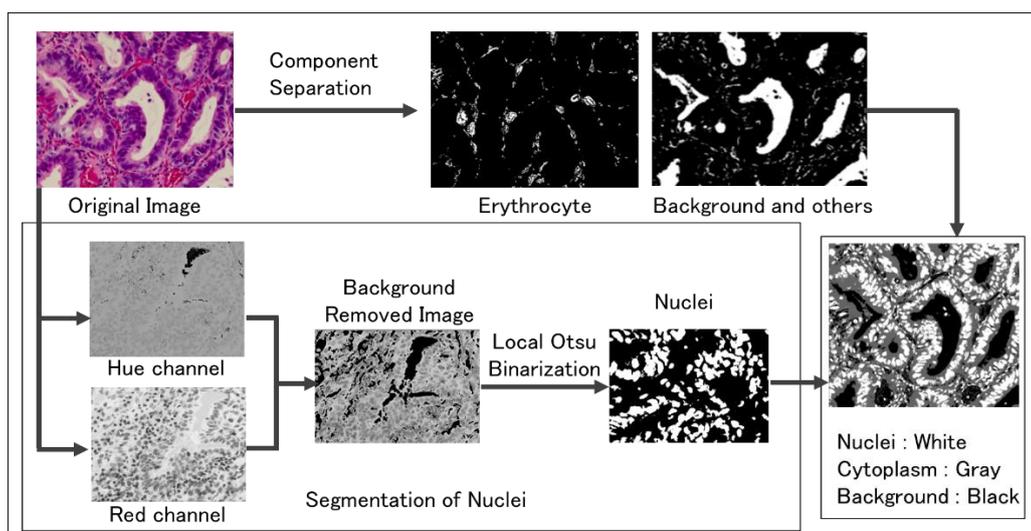


Figure 6.2.2 Generation method of 3-level images.

between the nucleus region and other areas in the R channel of RGB color space. H (Hue) channel out of HSV color space contains main properties of the type of color. It is defined technically as the qualitative aspect presented by visual color. Because the influence of uneven staining is restrained in H channel, H axis and R axis are employed to extract the nucleus region. The pixels which have obviously higher value both in R axis and H axis are accounted into nucleus region.

(1) Scaling

To prepare the nucleus region emphasized image, all pixels in the original image are calculated as the formula:

$$g(x) = \begin{cases} f_H(x) + k \times f_R(x), & (f_H(x) + k \times f_R(x) < 255) \\ 255, & (f_H(x) + k \times f_R(x) \geq 255) \end{cases} \quad (2)$$

where $g(x)$ is the nucleus region emphasized image. Where $f_H(x)$ and $f_R(x)$ represents H channel and R channel of the original image $f(x)$ correspondingly. Parameter k is determined from preliminary experiments.

2) Local Otsu Binarization

Otsu's method [65] is an adaptive threshold determining method by finding the minimal intra-class variance. Otsu's method is widely used in automatic image binarization in computer vision and image processing.

When using Otsu's method usually it is necessary to search for the threshold that minimizes the intra-class variance, which means the maximum variance between classes. Otsu's method operates directly on the gray level histogram $P(i)$:

$$\sigma_w^2(t) = q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t) \quad (3)$$

where the class probabilities are estimated as:

$$q_1(t) = \sum_{i=1}^t P(i) \quad (4)$$

$$q_2(t) = \sum_{i=t+1}^l P(i) \quad (5)$$

The class means are given by:

$$\mu_1(t) = \sum_{i=1}^t \frac{iP(i)}{q_1(t)} \quad (6)$$

$$\mu_2(t) = \sum_{i=t+1}^l \frac{iP(i)}{q_2(t)} \quad (7)$$

Finally, the individual class variances are:

$$\sigma_1^2(t) = \sum_{i=1}^t [i - \mu_1(t)]^2 \frac{P(i)}{q_1(t)} \quad (8)$$

$$\sigma_2^2(t) = \sum_{i=t+1}^l [i - \mu_2(t)]^2 \frac{P(i)}{q_2(t)} \quad (9)$$

Next, traverse the calculation and pick the value t that minimizes $\sigma_w^2(t)$.

Because the pathological image employed is large and staining is not even from part to part. It is not reasonable to binarize an image with the threshold calculated from the original Otsu's method in wide range. Thus, the proposed local Otsu binarization method to adapt the uneven staining of pathological image and this binarization method is free from image size. In the local Otsu binarization method, local uneven staining or artifact could be weakened without affecting other area.

In the proposed local Otsu binarization, the threshold of a pathological image λ was calculated using Otsu's method. Then segment the image into several small parts and compute threshold of each part λ_i . If the ratio of size of background area

is less than 50%, this part is binarized by threshold λ_i , else binarized by threshold λ .

After implementing the above scaling and local Otsu's method, binarized images consist of only nucleus can be generated.

(3) Segmentation of Background

The background regions of pathological image with high luminance level are expressed well in G channel out of RGB color space. The pixels with larger G value than the threshold which is determined by Otsu's method will be considered as background region. In background segmentation, because red blood cells neither belong to nuclear nor cytoplasm area, they are counted as the background. Therefore, red blood cells are segmented taken advantage in R channel in RGB color space according to the fact that red blood cells have a strong red component. In this way, binarized image includes background and others is generated.

(4) Image Integration

Two binary images are generated by adapting the above mentioned segmentations, (1) nucleus, (2) background and other areas. The region neither belongs to background area nor belongs to nucleus area is finally categorized as cytoplasm. After image integration, nucleus, background and cytoplasm areas are concluded in one image and possess three levels in order.

6.2.3 CILAC (Color Index Local Auto-Correlation)

CILAC was developed on the basis of Higher Local Auto-Correlation (HLAC), which has been demonstrated as a capable image descriptor and has been adopted in many image classification applications such as object counting [66], gesture recognition [67], video surveillance [68] and medical image classification [69, 70, 71]. HLAC is denoted as following.

$$X_N = \sum_{\mathbf{r}} I(\mathbf{r})I(\mathbf{r} + \mathbf{a}_1) \cdots I(\mathbf{r} + \mathbf{a}_N) \quad (10)$$

where I is the objective image, \mathbf{r} is the view position, and \mathbf{a}_i ($i = 1, 2, \dots, N$) is the displacement vector around \mathbf{r} , which indicates pixels to be calculated. Taking consideration of the balance of performance and computing cost, the proposed method adopts $N \in \{0, 1, 2\}$. HLAC features with different orders form a complete feature vector. Intuitively, HLAC feature are expressed by mask patterns

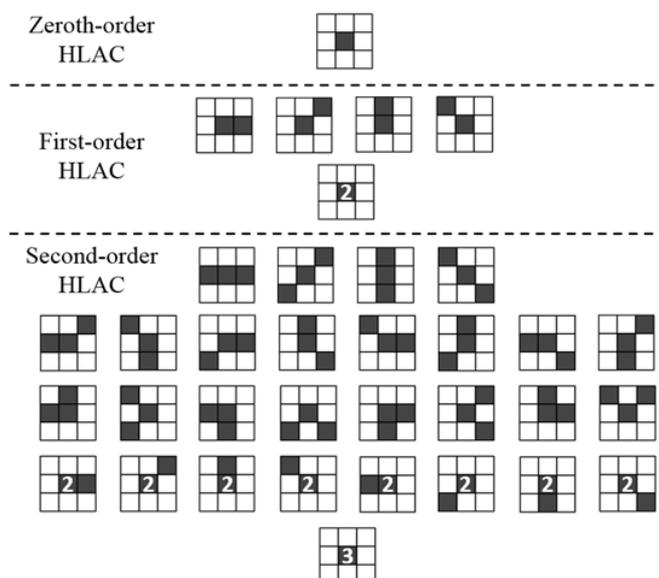


Figure 6.2.3 Gray-scale HLAC mask patterns.

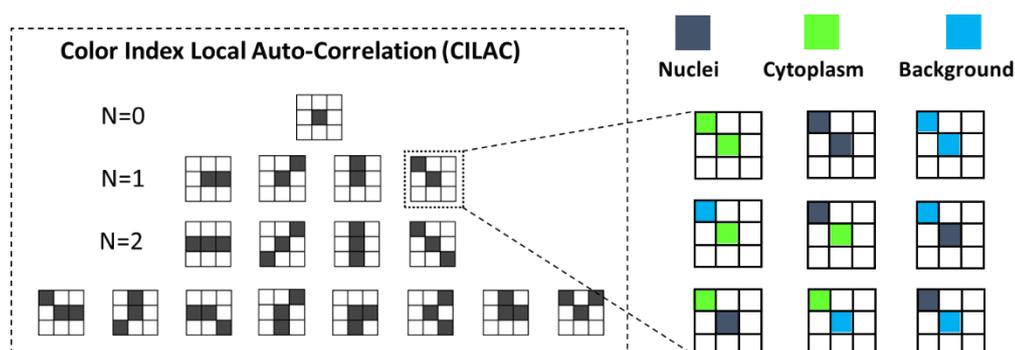


Figure 6.2.4 CILAC mask patterns.

as shown in Figure 6.2.3. Specifically, 0-order HLAC is used to evaluate the intensity of the reference pixel, while 1-order and 2-order HLAC is used to calculate auto-correlations between the reference pixel and surrounding correlative pixels, which indicate the edge and curvature, respectively. These morphological elements are deemed to be competent to cover the measurement requirement of cell-wise characteristics.

Taking the place of calculation of grayscale local-correlation, CILAC focuses on the specificity of co-occurrence neighboring pixels within the image. CILAC consists of a set of local patterns which are able to calculate both the local auto-correlations of different color levels and their statistical distribution. CILAC in order N ($N = 0, 1, 2$) is defined as below:

$$S_0(i) = \sum_r f_i(r) \quad (11)$$

$$S_1(i, j, a) = \sum_r f_i(r) f_j(r + a) \quad (12)$$

$$S_1(i, j, k, a, b) = \sum_r f_i(r) f_j(r + a) f_k(r + b) \quad (13)$$

Similar to HLAC, as S_N denotes N -order correlation. $f = \{e_1, e_2, e_3 \dots, e_D\}$ is a D -dimensional vector standing for D color indexes of an color-indexed image. r indicates the reference (central) pixel. a, b are different displacements of the surrounding inspected pixels, respectively. f_i, f_j and f_k denote the pixels taken into account corresponding to all displacements. In this paper, D is set to 3 according to three color indexes of the 3-level image. In that case, the 0th order CILAC ($N = 0$) draw out different color indexes themselves, and the 1st and 2nd order CILAC ($N = 1$ and $N = 2$) represent the local co-occurrences of different color indexes. Pathological components including nuclei, cytoplasm and background are expected synthetically vectorized by the CILAC patterns (Figure 6.2.4).

6.2.4 Dimensionality Reduction and Unsupervised Clustering

After CILAC feature extracted from 3-level pathology images. Principal component analysis (PCA, [72]) is also employed to reduce the dimensionality of feature vector space. Next, unsupervised K-means clustering [73] is used to separate images into several clusters within the feature vector space. Practically, in order to obtain clusters with large distance as possible, this paper sets the number of cluster $k=3$, and select the farthest two clusters in line with the visualized status within the coordinate space of finite principal components. Finally, the most distant two clusters are selected and assigned with +1 and -1. Pass through the above series of operations, the two clusters are available to be automatically generated and employed as “medium-level” training data for the 1st step fine-tuning.

6.3 Experiments and Discussion

(1) Experimental Procedures

In order to evaluate the effectiveness of our proposed stepwise knowledge transfer scheme using e automatically produced low-cost “medium-level” datasets based on CILAC, same as in Chapter 5, this dissertation makes use of three types

Table 6.3.1 Datasets used in experiments (cell-wise).

Data Type	Category	Training	Validation	Test
Medium-level Data	Cluster 1	5,016	558	-
	Cluster 2	3,949	439	-
High-level (Well-annotated) Data	Benign	5,400	1,620	2,700
	Malignant	5,400	1,620	2,700

of well-known deep neural networks, VGG-16, AlexNet and GoogLeNet (InceptionV3). With each of the deep neural networks, two separate procedures are conducted. (1) adopting fine-tuning only once with high-level well-annotated pathology images directly upon the model which has been pre-trained by low-level large-scale datasets (ImageNet). (2) adopting the 1st fine-tuning and 2nd fine-tuning in sequence with the “medium-level” data and high-level well-annotated pathology image data, respectively. Competitions are carried out between the two procedures based on the three deep neural networks stated above.

(2) Datasets

Same as Chapter 5, there are also three types of data utilized for the initialization (pre-training) of deep neural networks, the 1st stage knowledge transfer and the 2nd stage knowledge transfer, respectively. While ImageNet is still used as “low-level” data, well-annotated datasets including 5,400 benign and 5,400 malignant patches are used as “high-level” data, the composition of which is exactly the same as in the previous chapter. Meanwhile, as shown in 6.3.1, the validation dataset including 1,620 benign and 1,620 malignant patches and the test dataset of 2,700 benign and 2,700 malignant patches are employed again, without any overlap. As to the full-

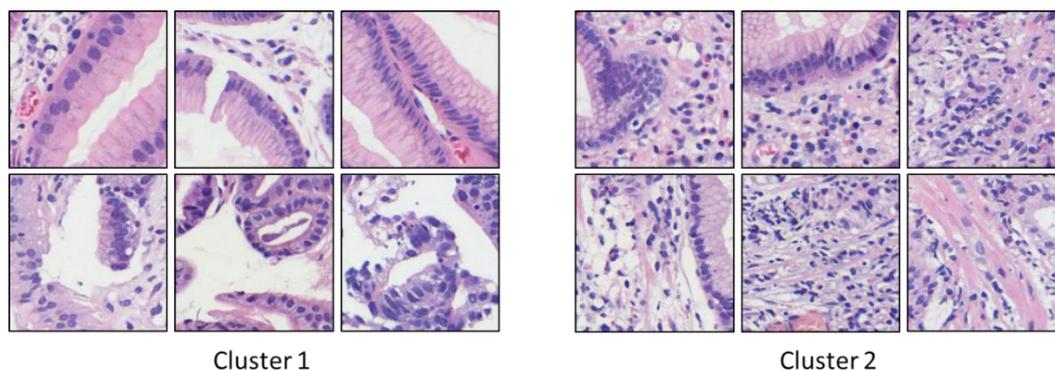


Figure 6.3.1 Example of automatically generated “Medium-level” data.

automated cell-wise “medium-level” data, by adopting unsupervised clustering upon more than 10,000 patches (256×256), it has been successful to obtain cluster 1 including 5,574 patches and 4,388 patches belong to cluster 2. Examples are depicted in Figure 6.3.1. Obviously, two clusters with different cell morphology were generated. Patches in cluster 1 seem to have larger but small nuclei, while patches in cluster 2 are small and dense. In the 1st-stage knowledge transfer, 90% of patches in each cluster are used for training, remaining 10% are used for validation. Validation data are completely separated from training data so that well-generalized model can be selected accordingly. All of these patches are cut off from whole pathology images without augmentation. It is noteworthy that there is no overlap between the “medium-level” datasets and the “high-level” datasets, and meanwhile no overlap among the training, validation and test datasets.

(3) Experimental Results

Results are presented and discussed among the rival performances of the regular one-stage knowledge transfer and the proposed stepwise knowledge transfer scheme. To be impersonal, this part also employs AUC, ACC, Precision and Recall as the evaluation criteria.

As denoted in Table 6.3.2, notably, in all of the couples of competitive schemes, our proposed two-step fine-tuning using “medium-level” dataset has yield reasonable improvement. Specifically, AUC value is raised by 0.021, 0.056 and 0.058, when CNN architectures VGG-16, AlexNet and Inception V3 are adopted respectively. Meanwhile, if we focus on ACC values, the fact is distinct that the greatest improvement happens when the proposed scheme using Inception V3 is adopted. The accuracy has remarkably increased from 0.779 to 0.865. Besides,

Table 6.3.2 Performances of the proposed two-stage knowledge transfer using cell-wise “medium-level” data.

Scheme	CNN Architecture											
	VGG-16				AlexNet				GoogLeNet (Inception V3)			
	AUC	ACC	Preci- sion	Recall	AUC	ACC	Preci- sion	Recall	AUC	ACC	Preci- sion	Recall
One-stage	0.936	0.836	0.96	0.70	0.867	0.794	0.80	0.79	0.881	0.779	0.79	0.78
Two-stage (Proposed)	0.957	0.873	0.87	0.87	0.923	0.845	0.85	0.84	0.939	0.865	0.87	0.86

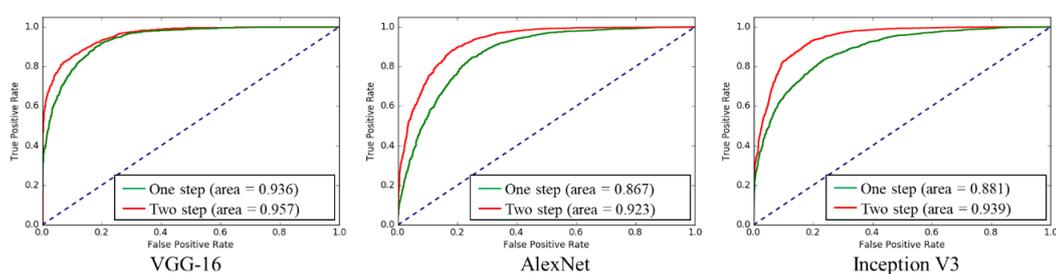


Figure 6.3.2 Performances of the proposed two-stage knowledge transfer using cell-wise “medium-level” data by ROC.

precision and recall, which are commonly used for medical image classification, are presenting similar trend to AUC and ACC. As more intuitively illustrated in Figure 6.3.2, three CNN architectures have produced three separate ROC Figures. The red curve denotes the two-stage knowledge transfer scheme using automatically generated “medium-level” dataset, while the green curve denotes the conventional one-stage knowledge transfer scheme. It is clear at a glance, in each figure, our proposed scheme possesses overwhelming area all along both the false positive rate axis and true positive rate axis. These results have illustrated that our proposed scheme is capable and rarely dependent on the deep neural network’s architecture and the amount of well-annotated data. Accordingly, it is well-founded that the proposed stepwise knowledge transfer scheme using cell-wise “medium-level” data automatically produced based on Color-Index Local Auto-Correlation (CILAC) has successfully boosted the performance of the pre-trained neural networks for gastric pathology image classification in various situations.

6.4 Summary

In this chapter, aiming to maximize the classification capacity of deep neural networks and alleviate the lack of annotated pathology data, a stepwise knowledge transfer scheme using based on automatic measurement with human understanding is developed. By extracting pathology-correlative information from unannotated pathology images with handcrafted feature CILAC, and making use of these materials as “medium-level” data to intermediately perform knowledge transfer to deep neural networks, it managed to make the deep neural networks acquire pathological knowledge following the way of human perception. By this means, the initial task and the final target task are expected to be bridged in a reasonable way. Meanwhile, compared to tissue-wise “medium-level” data, when there is a serious shortage of manpower, it is possible to generate cell-wise “medium-level” data for knowledge transfer at almost zero cost. In the experiments, the proposed scheme exerted adequate efficacy for boosting the classification performance and revealed high applicability for different CNN architectures.

Chapter 7

Stepwise Transfer of Multiple Domain Knowledge

7.1 Concept and Formation

In Chapter 5 and 6, the feasibility of transfer of pathological knowledge based on both direct human observation and automatic measurement with human understanding has been validated. These trials are all conform to the proposed conception that before make deep neural networks adaptive to benign/malignant classification task, it should understand the difference of general visual characteristics and fundamental pathology knowledges within the relational domains in line with usual perception manner and learning progress. After separately evaluating the two types of relational-domain knowledge, tissue-wise knowledge and cell-wise knowledge, this chapter will further discuss about the possibility of simultaneously using the two different knowledge to transfer in one scheme. Figure 7.1.1 shows the specific implementation procedures. Obviously, the steps of the relational domain are increased, which is equivalent to the further

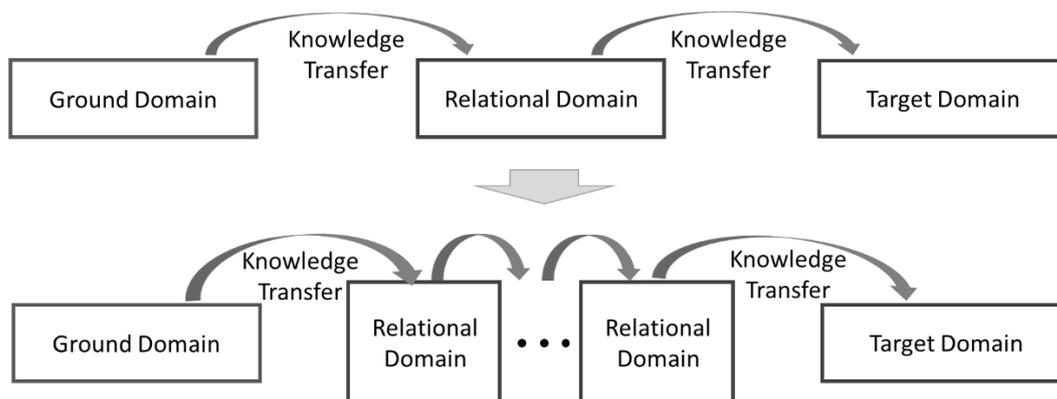


Figure 7.1.1 Procedures of Transfer of Multiple Domain Knowledge.

subdivision of the conversion process from the ground domain to the target domain. By increasing the level of relevance to the target domain, the specialization of deep neural networks is expected to be further strengthened. Additionally, for an existing question, that is the relationship between the amount of involved relational-domain data and the performance of the target domain, this part will also give the answer.

Furthermore, there is another the assumption that whether the source of knowledge to be transferred can be expanded. As shown in Figure 7.1.1, based on the competitive scheme already discussed, this paper will also introduce different kinds of pathological images in order to make a more comprehensive judgment on the scope of application and operability of the proposed knowledge transfer scheme for pathology image classification.

7.2 Stepwise Transfer of Multiple Domain Knowledge

This section will debate about the feasibility of transfer of multiple domain knowledge. Based on the flow shown in the lower part of Figure 7.1.1, the relational domain will be divided into two parts filled in order with the tissue-wise knowledge and cell-wise knowledge proposed in the previous chapters. Here, for the convenience of operation, the datasets used in Chapters 5 and 6 are directly employed again as shown in Table 7.2.1. The first part of relational domain, tissue-wise data, include three types, background, epithelium, and stroma, each with 15,000 training patches, while the second part of relational domain, cell-wise data, include two clusters, with 5,016 training patches and 3,949 training patches separately. As to the deep neural network, for an intuitive comparison this time only VGG-16 is adopted.

Table 7.2.1 Datasets used in experiments (multiple domain).

	Category	Training	Validation	Test
Relational Domain Part 1: Tissue-wise	Background	15,000	1,000	-
	Epithelium	15,000	1,000	-
	Stroma	15,000	1,000	-
Relational Domain Part 2: Cell-wise	Cluster 1	5016	558	-
	Cluster 2	3949	439	-
Target Domain	Benign	5400	1620	2700
	Malignant	5400	1620	2700

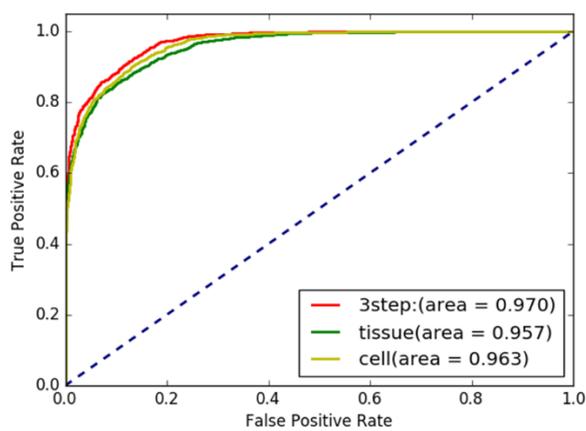


Figure 7.2.1 Results of transfer of multiple domain knowledge.

Results are depicted in Figure 7.2.1. Red curve, indicating the result of transfer of multiple domain knowledge, presents the topmost AUC value 0.970. It has substantially exceeded the results of the reference schemes using tissue-wise data and cell-wise data. This result provides evidence that the use of stepwise transfer of multiple domain knowledge is able to acquire more adaptability to improve the accuracy of the task in target domain, by introducing different levels of relational-domain knowledge.

However, one big concern comes out. In the experiments above, the data used in Chapter 5 and Chapter 6 were fully employed. Compared to the verification experiments for the two separate schemes, the amount of data introduced into the entire process of stepwise transfer has been accordingly increased. This makes it easy to suspect that the performance gains obtained in the above experiments were due to an increase in the amount of data rather than the use of stepwise transfer of multiple domain knowledge. Therefore, it is necessary to discuss about the factors of change. Here, an additional set of experiments are presented to address this

Table 7.2.2 Performances of schemes using datasets with fixed number of images.

Category	Tissue-wise	Cell-wise	Tissue-wise + Cell-wise
Number of images	8,000	8,000	4,000 + 4,000
AUC	0.961	0.965	0.969

answer. This time, the total amount of data introduced into the relational domain will be fixed, as shown in Table 7.2.2. In the first group, the tissue-wise data contains 4000 epithelium images and 4000 stroma images. In the second group, cell-wise data contains 8000 images covering two clusters. In the third group, 4000 tissue-wise images and 4000 cell-wise data images are employed respectively. In this way, since the total number of images in the three sets is consistent, the difference in performance caused by the difference in the amount of information will be largely suppressed.

From the results of this comparative experiment, it is easy to conclude that making use of two different data still brings a slight performance improvement when the amount of data introduced into the relational-domain is consistent. This also laterally provides a guarantee for the validity of the verification given in the previous experiments.

7.3 Extension of Relational Domain Knowledge

In order to further increase available intermediate data, this part will devise a knowledge transfer based on extension of relational-domain data, not limited to the same field as the target. Specifically, in all the descriptions so far, all images are collected from gastric pathology examination. This means that except for the initialization of the deep neural networks is implemented using ImageNet data, knowledge transfer corresponding to all the other domains is limited to the scope of the gastric pathology image. In the previous section it has been discussed about the proposed scheme of stepwise transfer of domain knowledge and its possible extensions in terms of process and form. However, if faced with more complicated actual situations, such as the imbalance of the amount of pathological image data in different parts (for example, the number of pathological images collected from certain organ is small, while the others are large), then only in the flow and form of the scheme, then the benefits of just extending the process and form of the scheme are greatly limited. In view of this situation, the following part proposes a scheme

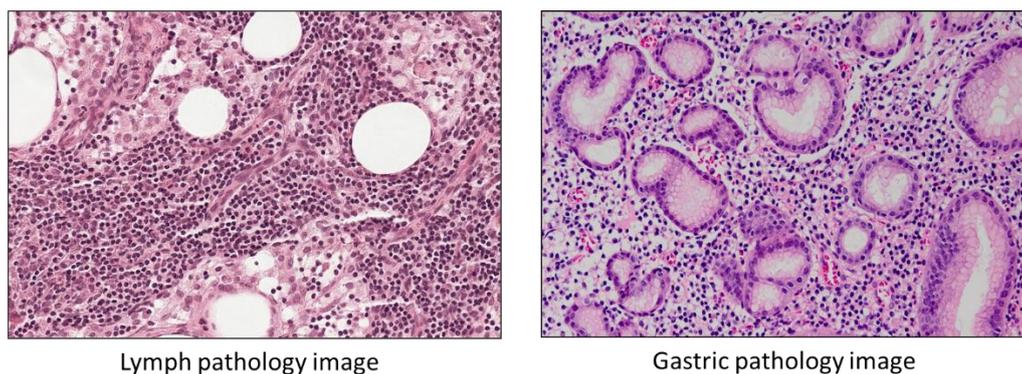


Figure 7.3.1 Different kinds of pathology images.

Table 7.3.1 Datasets including gastric pathology images and lymph pathology images within relational domain.

	Organ	Category	Training	Validation	Test
Relational Domain Part 1: Tissue-wise	Gastric	Background	15,000	1,000	-
		Epithelium	15,000	1,000	-
		Stroma	15,000	1,000	-
Relational Domain Part 2: Cell-wise	Gastric	Cluster 1	5,016	558	
		Cluster 2	3,949	439	
	Lymph	Cluster 1	3,954	440	-
		Cluster 2	1,822	203	-
Target Domain	Gastric	Benign	5,400	1,620	2,700
		Malignant	5,400	1,620	2,700

for performing knowledge transfer using different kinds of pathological images.

In the next experiment, cell-wise data generated using part of the open data of Camelyon 16 Challenge, in with a total of 400 breast cancer lymph node metastasis whole-slide images are contained. Obviously in Figure 7.3.1, although lymph pathology image and gastric pathology image both belong to pathology images, the presence or absence of glands, the distribution of cells and staining conditions are markedly different. The stepwise transfer of domain knowledge in this experiment will follow the pattern of the multiple domain proposed in the previous section. After initializing the deep neural network with ImageNet, the knowledge transfer within the relational-domain is first performed using tissue-wise data. After that, the knowledge transfer using cell-wise data will be divided into two cases. In the first case, fine-tuning is performed using only the gastric pathology image in the manner of 7.2. In the second case, the lymph pathology image was added to the gastric pathology image, and the pathological images of two different organs are

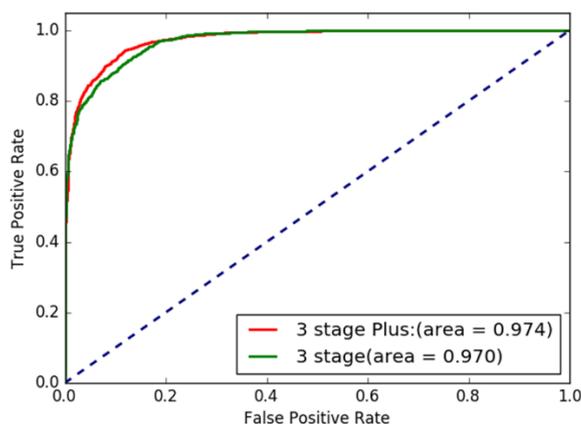


Figure 7.3.2 Performance of stepwise knowledge transfer scheme using pathology images of different organs.

combined into one dataset to fine-tuning the deep neural network. To compare the performance of recognition accuracy in the final task in both cases will help us to see if it is a feasible way to introduce pathology images of different organs into the proposed stepwise knowledge transfer scheme at the same time. According to performance of ROC, the red curve indicating the trial of extension of relational domain knowledge using both gastric and lymph images has achieved larger AUC than that using gastric images only.

Based on the above experimental results, it has been confirmed that the performance improvement are further achieved by stepwise knowledge transfer based on the extension of the relational-domain knowledge, when deep neural network managed to acquire more information contributing to the target domain. Therefore, when carrying out knowledge transfer based on extension of related domains, I believe that it is possible to make full use of data in common knowledge not only in the same organ but also in others.

Nevertheless, since the impact due to the increase in the amount of data must also be taken into account, in the way of 7.2, a set of additional experiments within which the numbers of data in the same level are fixed, will be implemented as a reference. This time, the data used for knowledge transfer will still be set to the order of ground-domain data (ImageNet), tissue-wise data, cell-wise data, target-domain data. Merely, in the cell-wise part, the first case will only use 4,000 gastric pathology images. Correspondingly, the second case would use a collection containing 2,000 gastric pathology images and 2,000 lymph pathology images. Thus, the total number of pathology images used in both cases is exactly the same.

Table 7.3.2 Performances of schemes using datasets with fixed number of images.

Category	Gastric cell-wise	Gastric + lymph cell-wise
Number of images	4,000	2,000 + 2,000
AUC	0.969	0.973

It can be seen from the results that the image obtained by using two different organs at the same time has a slightly higher final recognition performance than simply using the same image while keeping the total number of images unchanged. This result verifies the effectiveness of the previously mentioned experiments used to verify extended relational-domain knowledge, and on the other hand reflects the different types of pathological images sometimes in dealing with the classification of a pathological image. Instead, it will bring more helpful information. It is difficult for us to specifically examine which information has been introduced, and ultimately affect the target-domain recognition results. However, it is certain that some morphological features in different kinds of pathological images are common, and can help the specialized improvement of the deep learning network without any difference in a specific training environment.

7.4 Summary

In this chapter, my research further explored the extensibility that can be used for the proposed stepwise knowledge transfer scheme. The first exploration is about the form of transfer scheme. Under the framework of using the proposed knowledge transfer scheme including ground domain, relational domain, target domain. The artificially generated tissue-wise data based on direct observation and cell-wise data based on automatic measurement with human understanding are introduced into the knowledge transfer process in order. As the result, it is demonstrated superior to the simple scheme employing only one type of the knowledge. This exploration also confirms the hypothesis of "train deep neural networks in a human-understood manner and strengthening its adaptability for benign/malignant classification". The second exploration is about the types of pathology image used. In this part, cell-wise data is extended from the gastric pathology image to other kinds of images. In the experiment, lymph pathology images obtained from open data of Camelyon 16 Challenge are used. According to the experimental results, the use of cell-wise data mixed into the lymph pathology image also brings a small increase in the final

recognition performance. This result also provides suggestions for some special practical cases, such as when facing the imbalance of the amount of pathological image data in different organs. Actually, due to the disparity in the incidence of cancer in different parts, this problem may also be more important in the actual clinical diagnosis. Meanwhile, although this paper has discussed knowledge transfer scheme with up to three steps because of the experimental condition, the applicability of knowledge transfer with more steps may also be worthy of further discussion.

Chapter 8

Conclusions

This dissertation has presented the studies on stepwise transfer of domain knowledge for computer-aided diagnosis in pathology using deep neural networks. In order to alleviate the significant problem of insufficient well-annotated pathological image data when use deep learning based approaches to classify cancer images and normal images, the proposed scheme is developed based on transfer learning, which makes use of a pre-trained model for the task of another relational domain. On the premise of confirmation that knowledge gained from the ground domain is conducive to the performance improvement for pathology image classification, another issue has been realized that the knowledge transfer crossing domains with insufficient correlation probably lack of efficacy. Aiming to solve this issue, the proposed scheme put forward a stepwise processing flow including an intermediate domain which is related to but differing from the target-domain, thereby to bridge the knowledge domains far apart. In this way, deep neural networks are supposed to be evolved to a state more adaptable for the target task and finally able to achieve more competitive recognition accuracy of

benign/malignant classification, even though in the case of the limited amount of the well-annotated pathological image data for training.

To realize the proposed stepwise scheme, two types of knowledge to be transferred have been presented and involved in the form of “medium-level” data. One type can be generated directly based on human observation, while another type is generated based on automatic measurement with human understanding.

Specifically, the knowledge generated based on human observation is inspired by the distinct appearance among natural morphological status of pathology images. In this research, gastric pathology images are divided into three regions, background, stroma and epithelium, through simple annotation operations by non-professional manpower with only pathologist’s initial direction. The datasets are then utilized to establish the relational-domain tasks and realize the implementation of knowledge transfer. Comparative experiments have been performed among different deep neural networks and well-annotated pathology images of two sizes, it is noticeably found that the proposed two-stage knowledge transfer scheme is effective to contribute to boosting the classification accuracy in line with results in various situations.

Another type of knowledge to be transferred is generated based on automatic measurement with human understanding. By extracting pathology-correlative information from unannotated pathology images with handcrafted feature CILAC, and it was successful to use the feature vector space to separate images into several clusters, and employ these clusters as “medium-level” data to intermediately perform knowledge transfer to deep neural networks. Compared to tissue-wise “medium-level” data, when there is a serious shortage of manpower, it is possible to generate cell-wise “medium-level” data for knowledge transfer at almost zero cost. In the experiments, the stepwise scheme exerted adequate efficacy and applicability to various deep neural network models.

Furthermore, this research explored the extensibility that can be used for the proposed stepwise knowledge transfer scheme. The first exploration, which is related to the form of transfer scheme, introduced both tissue-wise data based on direct observation and cell-wise data based on automatic measurement with human understanding into the knowledge transfer process in order. As the result, the sequential knowledge transfer has brought further enhancement. In the second exploration, cell-wise data was extended from the gastric pathology image to other kinds of images. When lymph pathology images are concluded within the cell-wise

data together with gastric pathology image, performance of target task has shown slightly incensement. This result is considered meaningful to suggest for solutions of special practical cases when imbalance of the amount of pathological image data in different organs is faced.

Also, as this research has not reached, it is considered that detailed evaluation of the following elements is necessary when examining practical application based on the proposed scheme. (1) Relationship of the feature distribution and data amount between related domain and target domain. (2) Relationship between each layer of deep neural network and knowledge transferred (3) Optimization of hyper-parameters. In particular, I think that there are many parts that human have not figured out yet about the evidences that can show the relevance between the weights of each layer of the deep neural network and the acquired knowledge, hence, if there is found a method that can accurately measure changes around that, it is believed that the most advisable information can also be quantified and obtained.

In addition, with regard to the scope of application of the proposed scheme, all discussions in this research were conducted around pathology image recognition. Analogous to pathology images, other methods use deep learning in medical image recognition may also encounter problems such as insufficient data and data imbalance. In a similar situation, the stepwise knowledge transfer scheme proposed in this paper may be deemed as a clue to solve the problem. For example, when a mammography image classification is coped with a deep learning based method, the images having similar property such as general chest radiographic images may be introduced as useful knowledge to improve the adaptability of the deep learning models.

References

- [1] Ferlay J, Soerjomataram I, Ervik M et al., “GLOBOCAN 2012 v1.0,” Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11, 2013.
- [2] An Overview of The Yearly Total of Monthly Report of Population Dynamics Statistics In 2013, https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/kakutei13/dl/11_h7.pdf, available on 5 Jan. 2019.
- [3] American Cancer Society, Cancer Facts & Figures 2018, <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf>, available on 5 Jan. 2019.
- [4] Samita V Khadse, Ankur Aggarwal, Manjunath Sm et al., Biopsy as A Diagnostic Tool: A Review Introduction, Online Journal of BioSciences and Informatics, Vol: 5, Issue 3, 2013.
- [5] Kochi Medical Center, Introduction of Pathology Examination Department, <https://www2.khsc.or.jp/info/dtl.php?ID=1338&routekbn=S>, available on 5 Jan. 2019.
- [6] Anne Månsson, Mikael Adner, Ulf Höckerfelt et al., A distinct Toll-like receptor repertoire in human tonsillar B cells, directly activated by Pam3CSK4, R-837 and CpG-2006 stimulation, Immunology, Volume 118, Issue 4, pp 539-548, 2006.
- [7] Peter N. Furness, The Use of Digital Images in Pathology, Pathology, Volume

- 183, Issue 3, pp 253-263, 1997.
- [8] A. Afework, M. D. Beynon, F. Bustamante et al., Digital dynamic telepathology--the Virtual Microscope, Proc. AMIA Symposium, pp 912-916, 1998.
- [9] Stanley J. Robboy, Sally Weintraub, Andrew E. Horvath et al., "Pathologist Workforce in the United States," Archives of Pathology & Laboratory Medicine, vol. 137, no. 12, pp. 1723-1732, 2013.
- [10] The Japanese Society of Pathology Guideline 2015, The Japanese Society of Pathology, pp. 6, 2015.
- [11] How Telemedicine Answers Global Pathology Demands, <https://proscia.com/blog/2015/07/14/global-crisis-digital-solution>, available on 19 Mar. 2018.
- [12] The Japanese Society of Pathology, List of Authorized Pathologists, <http://pathology.or.jp/senmoni/board-certified.html>, available on 5 Jan. 2019.
- [13] Michel Jondet, Régis Agoli-Agbo and Louis Dehennin, "Automatic Measurement of Epithelium Differentiation and Classification of Cervical Intraepithelial Neoplasia by Computerized Image Analysis," Diagnostic Pathology, vol. 5, no. 1:7, 2010.
- [14] Abdelrahim N. Esgiar, R. Naguib and Bayan S. Sharif et al., "Microscopic Image Analysis for Quantitative Measurement and Feature Identification of Normal and Cancerous Colonic Mucosa," IEEE Transactions on Information Technology in Biomedicine, vol. 2, no. 3, pp. 197-203, 1998.
- [15] J. Diamond, N. H. Anderson, P. H. Bartels et al., "The Use of Morphological Characteristics and Texture Analysis in the Identification of Tissue Composition in Prostatic Neoplasia," Human Pathology, vol. 35, no. 9, pp. 1121-1131, 2004.
- [16] K. Masood and N. Rajpoot, "Texture Based Classification of Hyperspectral Colon Biopsy Samples Using CLBP," International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1011-1014, 2009.
- [17] O. Sertel, J. Kong, H. Shimada et al., "Computer-aided Prognosis of Neuroblastoma on Whole-slide Images: Classification of Stromal Development," Pattern Recognition, vol. 42, no. 6, pp. 1093-1103, 2009.
- [18] Jakob N. Kather, Cleo-Aron Weis, Francesco Bianconi et al., "Multi-class Texture Analysis in Colorectal Cancer Histology," Scientific Reports, 6:27988, 2016.

- [19] J. Qu, H. Nosato, H. Sakanashi et al., “Cancer Detection From Pathological Images Using Higher-order Local Autocorrelation,” IEEE 2012 11th International Conference on Signal Processing, pp. 1198-1201, 2012.
- [20] Mohammad Peikari, Judit Zubovits, Gina Clarke et al., “Clustering Analysis for Semi-supervised Learning Improves Classification Performance of Digital Pathology,” International Conference On Medical Image Computing & Computer Assisted Intervention 2015: Machine Learning in Medical Imaging, pp. 263-270, 2015.
- [21] A. Tabesh, M. Teverovskiy, H.Y. Pang, et al., “Multifeature Prostate Cancer Diagnosis and Gleason Grading of Histological Images,” IEEE Transactions on Medical Imaging, vol. 26, no. 10, pp. 1366-1378, 2007.
- [22] K. Nguyen, A. Sarkar and A. K. Jain. “Structure and Context in Prostatic Gland Segmentation and Classification,” International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 115-123, 2012.
- [23] Po-Whei Huang and Cheng-Hsiung Lee, “Automatic Classification for Pathological Prostate Images Based on Fractal Analysis,” IEEE Transactions on Medical Imaging, vol. 28, no. 7, pp. 1037-1050, 2009.
- [24] Jia Qu, Hirokazu Nosato, Hidenori Sakanashi et al., “Computational Cancer Detection of Pathological Images Based on An Optimization Method For Color-Index Local Auto-Correlation Feature Extraction,” IEEE 11th International Symposium on Biomedical Imaging, pp. 822-825, 2014.
- [25] Korsuk Sirinukunwattana, David Snead and Nasir M. Rajpoot, “A Novel Texture Descriptor for Detection of Glandular Structures in Colon Histology Images,” SPIE Medical Imaging 2015: Digital Pathology, 94200S, 2015.
- [26] Dinggang Shen, Guorong Wu and Heung-Il Suk, “Deep Learning in Medical Image Analysis,” Annual Review of Biomedical Engineering, no. 19, pp. 221–248. 2017.
- [27] Raphaël Marée, “The Need for Careful Data Collection for Pattern Recognition in Digital Pathology,” Journal of Pathology Informatics. vol. 8, no. 19, 2017.
- [28] Hao Chen, Xiaojuan Qi and Lequan Yu, “DCAN: Deep Contour-Aware Networks for Accurate Gland Segmentation,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 2487-2496, 2016.
- [29] B. Bejnordi, G. Litjens, N. Timofeeva et al., “Stain Specific Standardization of Whole-Slide Histopathological Images,” IEEE Transactions on Medical

- Imaging, vol. 35, issue 2, pp. 404-415, 2016.
- [30] A. Krizhevsky, I. Sutskever and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," 25th International Conference on Neural Information Processing Systems, vol. 1, pp. 1097-1105, 2012.
- [31] K. Fukushima and S. Miyake, "Neocognitron: A New Algorithm for Pattern Recognition Tolerant of Deformations and Shifts in Position," Pattern Recognition, vol. 15, no. 6, pp. 455-469, 1982.
- [32] LeCun, Y., Bengio, Y., Convolutional networks for images, speech, and time series, The handbook of brain theory and neural networks, 3361(10), 1995.
- [33] Andrew Janowczyk and Anant Madabhushi, "Deep Learning for Digital Pathology Image Analysis: A Comprehensive Tutorial With Selected Use Cases," Journal of Pathology Informatics, vol. 7, no. 29, 2016.
- [34] K. He, X. Zhang, S. Ren et al., "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.
- [35] Thibaut Durand, Taylor Mordan, Nicolas Thome et al., "WILDCAT: Weakly Supervised Learning of Deep ConvNets for Image Classification, Pointwise Localization and Segmentation," IEEE Conference on Computer Vision and Pattern Recognition, pp. 642-651, 2017.
- [36] K Simonyan and A Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," International Conference on Learning Representations, 2015.
- [37] N. Dhungel, G. Carneiro and A. P. Bradley, "Deep Learning and Structured Prediction for the Segmentation of Mass in Mammograms," International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 605-612, 2015.
- [38] Zizhao Zhang, Yuanpu Xie, Fuyong Xing et al., "MDNet: A Semantically and Visually Interpretable Medical Image Diagnosis Network," IEEE Conference on Computer Vision and Pattern Recognition, pp. 6428-6436, 2017.
- [39] Mohammad Havaei, Axel Davy, David W. Farley et al., "Brain Tumor Segmentation with Deep Neural Networks," Medical Image Analysis, vol. 35, pp. 18-31, 2017.
- [40] Gustavo Carneiro, Jacinto C. Nascimento and António Freitas, "The Segmentation of the Left Ventricle of the Heart From Ultrasound Data Using Deep Learning Architectures and Derivative-Based Search Methods," IEEE

- Transactions on Image Processing, vol. 21, no. 3, pp. 968-982, 2012.
- [41] Y. Yuan, M. Q. Meng, “Deep Learning for Polyp Recognition in Wireless Capsule Endoscopy Images,” *Medical Physics*, vol. 44, no. 4, pp. 1379-1389, 2017.
- [42] Yan Xu, Zhipeng Jia, LiangBo Wang et al., “Large Scale Tissue Histopathology Image Classification, Segmentation, and Visualization via Deep Convolutional Activation Features,” *BMC Bioinformatics*, 18:281, 2017.
- [43] L. Hou, D. Samaras and TM. Kurc, “Patch-based Convolutional Neural Network for Whole Slide Tissue Image Classification,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2424-2433, 2016.
- [44] Xu J, Luo X, Wang G et al., “A Deep Convolutional Neural Network for Segmenting and Classifying Epithelial and Stromal Regions in Histopathological Images,” *Neurocomputing*, vol. 191, pp. 214-223, 2016.
- [45] Korsuk Sirinukunwattana, Josien P. W. Pluim, Hao Chen et al. “Gland Segmentation in Colon Histology Images: The GlaS Challenge Contest,” *Medical Image Analysis*, vol. 35, pp. 489-502, 2017.
- [46] Zhongyi Han, Benzhenq Wei, Yuanjie Zheng et al., “Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model,” *Scientific Reports*, no. 4172, 2017.
- [47] Geert Litjens, Peter Bandi, Babak Ehteshami Bejnordi et al., 1399 H&E-Stained Sentinel Lymph Node Sections of Breast Cancer Patients: The CAMELYON Dataset, *GigaScience*, Volume 7, Issue 6, 1 June 2018.
- [48] Dayong Wang, Aditya Khosla, Rishab Gargeya et al., *Deep Learning for Identifying Metastatic Breast Cancer*, arXiv:1606.05718.
- [49] Kaiming He, Xiangyu Zhang, Shaoqing Ren, *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*, *The IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026-1034.
- [50] F. Ciompi, O. Gessink, B. E. Bejnordi et al., “The Importance of Stain Normalization in Colorectal Tissue Classification with Convolutional Networks,” *IEEE International Symposium in Biomedical Imaging*, 2017.
- [51] Manan Shah, Christopher Rubadue, David Suster et al., “Deep Learning Assessment of Tumor Proliferation in Breast Cancer Histological Images,” arXiv:1610.03467.
- [52] Junhua Mao, Wei Xu, Yi Yang et al., *Deep Captioning with Multimodal*

- Recurrent Neural Networks (m-RNN), arXiv:1412.6632.
- [53] Anselm Brachmann, Erhardt Barth and Christoph Redies, 2017, Using CNN Features to Better Understand What Makes Visual Artworks Special, *Frontiers in Psychology*, 2017; 8: 830.
- [54] H. Chen, Q. Dou, D. Ni et al., 2015, Automatic Fetal Ultrasound Standard Plane Detection using Knowledge Transferred Recurrent Neural Networks, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 507-514.
- [55] Hoo Chang Shin, Holger R. Roth, Mingchen Gao et al., 2016, Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning, *IEEE Transactions on Medical Imaging*, vol. 35, issue 5, pp. 1285-1298.
- [56] J. Yosinski, J. Clune, Y. Bengio et al., 2014, How Transferable Are Features in Deep Neural Networks? *Annual Conference on Neural Information Processing Systems*, pp. 3320–3328.
- [57] Aiga Suzuki, Satoshi Suzuki, Shoji Kido et al., “A 2-staged Transfer Learning Method with Deep Convolutional Neural Network for Diffuse Lung Disease Analysis,” *Proc. of the 2017 Intl. Forum on Medical Imaging in Asia*, pp. 160-163, 2017.
- [58] Jia Qu, Nobuyuki Hiruta, Kensuke Terai, Hirokazu Nosato, Masahiro Murakawa, and Hidenori Sakanashi, “Gastric Pathology Image Classification Using Stepwise Fine-Tuning for Deep Neural Networks”, *Journal of Healthcare Engineering*, Volume 2018, Article ID 8961781.
- [59] M. Guarino, P. Micheli and F. Pallotti, “Pathological Relevance of Epithelial and Mesenchymal Phenotype Plasticity,” *Pathology - Research and Practice*, vol. 195, no. 6, pp. 379-89, 1999.
- [60] Bryony S. Wiseman and Zena Werb, “Stromal Effects on Mammary Gland Development and Breast Cancer”, *Science* vol. 296, no. 5570, pp. 1046-1049, 2002.
- [61] Olga Russakovsky, Jia Deng, Hao Su et al., “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision*, vol. 115, issue 3, pp 211-252, 2015.
- [62] M. Sokolova and G. Lapalme, “A Systematic Analysis of Performance Measures for Classification Tasks,” *Information Processing & Management*, vol. 45, no. 4, pp. 427-437, 2009.

- [63] Jia Qu, Hirokazu Nosato, Hidenori Sakanashi et al., 2014, Computational Cancer Detection of Pathological Images Based on An Optimization Method For Color-Index Local Auto-Correlation Feature Extraction, IEEE 11th International Symposium on Biomedical Imaging, pp. 822-825.
- [64] T. Kobayashi and N. Otsu, 2009, Color Image Feature Extraction Using Color Index Local Auto-Correlations, International Conference on Acoustics, Speech, and Signal Processing, pp. 1057-1060.
- [65] Nobuyuki Otsu, A threshold selection method from gray-level histograms, IEEE transactions on systems, man, and cybernetics, Vol. SMC-9, No. 1, 1979.
- [66] T. Kobayashi, T. Hosaka, S. Mimura, T. Hayashi and N. Otsu, "HLAC Approach to Automatic Object Counting," Proceeding of Bio-inspired, Learning and Intelligent Systems for Security, 40-45, 2008.
- [67] Takio Kurita and Satoru Hayamizu, "Gesture Recognition Using HLAC Features of PARCOR Images", IEICE Transactions on Information and Systems E86-D(4): 719-726, 2003.
- [68] S. Mimura, K. Itoh, T. Kobayashi, T. Takigawa, A. Tajima, A. Sawamura and N. Otsu, "The Cow Gait Recognition Using CHLAC," Proceeding of Bio-inspired, Learning and Intelligent Systems for Security, 56-57, 2008.
- [69] J. Qu, H. Nosato, H. Sakanashi, K. Terai and N. Hiruta, "Cancer Detection from Pathological Images Using Higher-Order Local Autocorrelation Feature," Proceeding of IEEE International Conference on Signal Processing, 1198-1201, 2012.
- [70] Erzhong Hu, Hirokazu Nosato, Hidenori Sakanashi and Masahiro Murakawa, "A Modified Anomaly Detection Method for Capsule Endoscopy Images Using Non-Linear Color Conversion and Higher-Order Local Auto-Correlation (HLAC) ", Proc. of IEEE Int. Conf. Engineering in Medicine and Biology Society (EMBC 2013), pp.5477-5480, Jul. 2013.
- [71] Hirokazu Nosato, Hidenori Sakanashi, Eiichi Takahashi et al., An Objective Evaluation Method of Ulcerative Colitis with Optical Colonoscopy Images Based on Higher Order Local Auto-Correlation Features.
- [72] B. Moore, Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction, IEEE Transactions on Automatic Control Volume 26 , Issue 1 , February 1981.
- [73] MacQueen, J. B., Some Methods for Classification and Analysis of Multivariate Observations", Proceedings of 5th Berkeley Symposium on

Mathematical Statistics and Probability. University of California Press. pp. 281–297, 1967.

Ethical approvals

The pathology images used in this study were provided from Sakura Medical Center (provider) to the National Institute of Advanced Industrial Science and Technology (the place where this study is carried out), subjected to appropriate procedures under the research plan approved by the ethical review of Sakura Medical Center and the National Institute of Advanced Industrial Science and Technology. All the images were provided after being randomly extracted and anonymized from histopathological specimens already diagnosed, after guaranteeing the opportunity to disclose and reject information to the research subjects.

List of publications

1. 野里 博和, 岩田 健司, 坂無 英徳, 高橋 栄一, 曲 佳, “病理組織診断支援のための高次局所自己相関特徴に基づく画像認識手法”, 第 10 回日本テレパソロジー・バーチャルマイクロスコープ研究会, 京都, 2011.
2. Jia Qu, Hirokazu Nosato, Hidenori Sakanashi, “Study on segmentation of pathological images for anomaly detection”, 情報処理学会第 74 回全国大会, 名古屋, 2012.
3. Qu J., Nosato H., Sakanashi H., Terai K., Hiruta N., “Cancer Detection From Pathological Images Using Higher-order Local Autocorrelation”, IEEE 2012 11th International Conference on Signal Processing (ICSP 2012), Beijing, Oct. 2012.
4. 野里 博和, 曲 佳, 坂無 英徳, 高橋 栄一, 村川 正宏, 寺井 謙介, 蛭田 啓之, “デジタル病理組織画像のための画像認識手法”, 第 11 回日本テレパソロジー・バーチャルマイクロスコープ研究会, 沖縄, 2012.
5. 曲 佳, 坂無 英徳, 野里 博和, 高橋 栄一, 寺井 謙介, 蛭田啓之, カラーテクスチャー解析によるデジタル病理組織画像からの異常検出, 第 12 回日本デジタルパソロジー研究会総会, 東京, 2013.

6. 野里 博和, 曲 佳, 坂無 英徳, 高橋 栄一, 村川 正宏, 寺井 謙介, 蛭田 啓之, カラーテクスチャー解析によるバーチャルスライド画像からの異常検出, 第 13 回日本テレパソロジー・バーチャルマイクロコピー研究会総会, 青森, 2014.
7. Jia Qu, Hirokazu Nosato, Hidenori Sakanashi, Eiichi Takahashi, Kensuke Terai, Nobuyuki Hiruta, “Computational cancer detection of pathological images based on an optimization method for color-index local auto-correlation feature extraction”, IEEE 2014 11th International Symposium on Biomedical Imaging (ISBI 2014), Beijing, Apr. 2014
8. 坂無 英徳, 曲 佳, 野里 博和, 高橋 栄一, 寺井 謙介, 蛭田啓之, Whole Slide Imaging データにおける正常状態の定量化と異常検出, 第 14 回日本デジタルパソロジー研究会総会, 倉敷, 2015.
9. Jia Qu, Nobuyuki Hiruta, Kensuke Terai, Hirokazu Nosato, Masahiro Murakawa, and Hidenori Sakanashi, “Gastric Pathology Image Classification Using Stepwise Fine-Tuning for Deep Neural Networks” , Journal of Healthcare Engineering, Volume 2018, Article ID 8961781.
10. Jia Qu, Hirokazu Nosato, Masahiro Murakawa, Kensuke Terai, Nobuyuki Hiruta, Hidenori Sakanashi, “Cancer detection for pathology images using stepwise fine-tuned deep neural networks” , 第 17 回日本デジタルパソロジー研究会総会, 呉, 2018.
11. Jia Qu, Hirokazu Nosato, Masahiro Murakawa, Kensuke Terai, Nobuyuki Hiruta, Hidenori Sakanashi, “Enhanced deep learning for pathology image classification: A knowledge transfer based stepwise fine-tuning scheme”, BIOIMAGING 2019, Prague, 2019. (Accepted)

Acknowledgements

I would like to send my most sincere gratitude to my supervisor Prof. Hidenori Sakanashi and Prof. Masahiro Murakawa. Many years ago, when I was coming to study in Japan and in the face of a totally unfamiliar environment, you encouraged me and brought me into the world to explore new technologies. Every progress I have made over the years has been inseparable from your teachings. At the same time, you have also set an example of a good and rigorous researcher in my heart.

I would also like to thank Prof. Koichi Mizutani for the special thanks. As my supervisor in University of Tsukuba, Prof. Mizutani always cares about my academic situation. He gave me a firm courage when I was hesitant and embarrassed during a period. He also put forward a lot of very valuable and constructive suggestions for my research.

My research has also been greatly assisted by Prof. Hirokazu Nosato. He gave countless valuable suggestions for my research and his excellent research ability has taught me a lot.

My research also benefited a lot from M.D. Nobuyuki Hiruta and M.D. Kensuke Terai from Toho University Sakura Medical Center. In their busy schedule, they not only provided valuable experimental data for my research, but also provided very professional advice and guidance for my research content.

I would also like to thank Prof. Yasuyo Kita of the degree review board. Her pertinent suggestions in the review are believed able to help my research achieve a

higher level of completion.

My thanks are also to all the past and present members of the Artificial Intelligence Applications Research Team at AIST. They have given me great support in all aspects of my research life. My progress cannot be separated from your help.

Last but not the least, I would also like to thank my beloved family. They give me the strength and courage to keep me going to complete the research. Because of you, my life is always full of meaning and color.