

東京大学  
情報理工学系研究科 創造情報学専攻  
修士論文

**病理画像解析支援のための深層能動学習に関する研究**  
Cost effective learning of pathological images with deep active learning

**福田 圭佑**  
Keisuke Fukuta

**指導教員 原田 達也 教授**

2018 年 1 月



# 概要

病理標本を撮像して診断を行うデジタルパソロジー技術の普及に伴い、画像認識技術によって病理医の診断を補助するための研究が数多く行われている。一方、高精度な識別率を達成する学習モデルを獲得するには、大規模なラベル付きデータを構築する必要がある。病理画像等の医療画像の学習データセットの作成には高度に専門的な知識が不可欠であるため、医師にとって大きな負担となってしまうアノテーションコストが非常に高い。そこで本研究では、画像群の中からモデルの識別精度向上に寄与する可能性の高いサンプルのみにアノテーションを付与することでアノテーションコストを抑えつつ高精度なモデルを獲得する能動学習の枠組みを採用する。また、能動学習を病理画像に適用するための工夫だけでなく、近年の画像認識分野で目覚ましい成果を挙げている Convolutional Neural Network (CNN) を能動学習に適用するための新たな手法を提案し、大規模病理画像データセットに対して実験を行い有効性を検証した。



# Abstract

Along with the spread of digital pathology technology, many researches have been conducted to assist diagnosis of a pathologist using computational image analysis. On the other hand, it is often necessary to construct large-scale labeled dataset to acquire a model that achieves high classification accuracy. Since training datasets in medical domain such as pathological images require highly specialized knowledge, it becomes a heavy burden to the doctor and the annotation cost is very high. Therefore, in this research, we adopt active learning framework to alleviate this issue without sacrificing high accuracy. Active Learning is a framework which annotates only the informative samples from an unlabeled image sets which are likely to improve a model accuracy. In addition to propose a method to apply active learning to pathological images, we propose a new method for applying Convolutional Neural Network (CNN) to active learning. We experimented on the large scale pathological image datasets and verified its effectiveness.



# 目次

<b>第 1 章</b>	<b>序論</b>	<b>1</b>
1.1	本研究の背景 . . . . .	1
1.2	本研究の目的 . . . . .	2
1.3	本研究の構成 . . . . .	2
1.4	本研究の貢献 . . . . .	2
<b>第 2 章</b>	<b>関連研究</b>	<b>4</b>
2.1	画像認識における深層学習 . . . . .	4
2.2	病理画像解析 . . . . .	6
2.3	能動学習 . . . . .	7
2.4	関連する先行研究と本研究の位置づけ . . . . .	11
2.5	まとめ . . . . .	12
<b>第 3 章</b>	<b>提案手法</b>	<b>13</b>
3.1	概要 . . . . .	13
3.2	Query by Dropout predictions . . . . .	14
3.3	推論時での Data Augmentation の利用 . . . . .	15
3.4	提案手法の動作原理 . . . . .	15
3.5	バッチ型能動学習への拡張 . . . . .	15
3.6	アルゴリズムの詳細 . . . . .	16
<b>第 4 章</b>	<b>実験 1 : MNIST を用いた予備実験</b>	<b>18</b>
4.1	概要 . . . . .	18
4.2	実験設定 . . . . .	18
4.3	実験結果 . . . . .	20
4.4	考察 . . . . .	20
<b>第 5 章</b>	<b>実験 2 : 病理画像データセットを用いた実験</b>	<b>22</b>
5.1	概要 . . . . .	22
5.2	実験設定 . . . . .	22

vi 目次

5.3	実験結果 . . . . .	23
5.4	考察 . . . . .	23
<b>第 6 章</b>	<b>結論</b>	<b>24</b>
6.1	結論 . . . . .	24
6.2	今後の課題 . . . . .	24
<b>参考文献</b>		<b>26</b>



# 図目次

2.1	Dropout の模式図。(左) 通常のネットワーク (右) いくつかのニューロンを Drop されたネットワーク . . . . .	5
2.2	病理画像の例。(ToDo: もうちょいいい感じのやつ) . . . . .	6
3.1	本研究で提案するシステムの概念図 . . . . .	14
3.2	Dropout によってサンプリングされた部分ネットワークによって Committee を形成する。 . . . .	15
4.1	各手法を利用した場合のラベル付きサンプル数の増加に対するテスト精度の変化を示した図 . . . . .	20
5.1	各手法を利用した場合のラベル付きサンプル数の増加に対するテスト精度の変化を示した図 . . . . .	23



# 表目次

3.1	比較実験の結果 . . . . .	17
4.1	MNIST の実験に使用した CNN の構造を示す。 . . . .	18
4.2	それぞれのクエリ選考基準を利用した場合にテスト精度を達成するために要したラベルの数 . . . . .	21



# 第 1 章

## 序論

### 1.1 本研究の背景

病理診断は、病理医がスライドガラス上の標本を観察することにより行われる。近年では標本全体のデジタル画像を撮影し、ディスプレイに表示することで細胞の組織を観察して診断を行うデジタルパソロジーの普及が進んでいる [1]。このデジタル画像は Whole slide image, WSI と呼ばれる。WSI は 1 スライド当たりギガピクセルに及ぶ巨大な高解像画像で、組織すべてを隅々まで確認して異常がないか診断を行うのは熟練の病理医でも時間を要する。

一方、デジタルパソロジーの普及に伴って WSI が大量に保管されるようになり、機械学習を利用した画像解析技術の応用によって病理医の診断を補助するための研究が数多く行われている [2, 3, 4]。中でも、WSI を入力として対応する病名を出力する自動診断に関する研究が盛んに行われている [5, 6]。

しかし、近年の画像認識分野において目覚ましい成果を挙げている Convolutional Neural Network (CNN) をはじめとするように、機械学習手法を用い、汎化性能の高い識別器を構築するには大量のラベル付きデータが必要となることが多い。病理画像などの医療画像の学習データセットの作成には高度に専門的な知識が不可欠であるため、医師への大きな負担となり、アノテーションコストが非常に高い。また、特に病理画像においては、1 枚ギガピクセルに及ぶ WSI に対する詳細なアノテーションの付与には莫大なコストが伴う。Camelyon Dataset とは、Camelyon Grand Challenge[7] にて公開されたデータセットで、1000 枚にも及ぶ WSI に対して乳癌のリンパ節転移の存在する領域への詳細なアノテーションが付与されている。WSI 1 枚のアノテーションにつき熟練の病理医が 1 時間かけて行ったとされ、延べ 1000 時間をこのデータセット作成に費やしたという報告もある。

教師データ作成コストが問題となるのは医療画像解析に限った話ではなく、機械学習では少量のデータから効率的に学習する枠組みの研究も盛んに行われている。アノテーションコストが特に高い状況に対する機械学習のアプローチとして、能動学習と呼ばれる枠組みがある。

教師付き学習は入力（質問）と出力（答え）の組からなる訓練データを用いて、その背後に潜んでいる入出力関係（関数）を学習する問題であり、しばしば問題集と解答集が与えられてテスト問題でできるだけ高い点数を取る状況に例えられる。またその例えでは、半教師付き学

## 2 第1章 序論

習は問題集と一部の解答のみが与えられる状況だと言える。ここで、能動学習とは、教師なし学習と同様に問題集のみが与えられるが、学習の過程でわからないところを教師に聞くことができる枠組みであると例えることが出来る。つまり、大量に与えられたラベルなしデータの中から、モデルの識別精度向上に寄与する可能性の高いサンプル（クエリ）を選択し、オラクル（アノテーター、もしくはドメインの専門家）にアノテーションをその都度付与してもらうことで、アノテーションコストを最小化しつつモデルの識別率を向上させるアプローチである。クエリの選考基準における理論的な研究だけでなく、実用的なアプリケーションに関する研究も近年増加しており、(ToDo: 引用) 高いアノテーションコストが大きな課題である病理画像解析への適用親和性は非常に高いと言える。

しかしそれらの研究は、一部の例外を除き高次元のデータへの scalability に欠く物が多く、計算コストの面から CNN のような巨大なパラメータを持つモデルを利用できない場合が多い。また、能動学習の多くでは特徴量が変わらないという状況での問題設定を置いていることが多く、表現学習を同時に行う深層学習を用いた研究は少ない。(ToDo: ここにもう一言書いてもいいかも)

### 1.2 本研究の目的

本研究では、病理画像解析において病理医が支払うコストを最小にしつつモデルの識別精度を担保するためのアプローチとして、ラベル付き学習データを学習の過程で機械とアノテーターとの interaction を通じて増加させていく能動学習を採用したシステムの構築を目的とする。能動学習を病理画像に適用するための工夫のみならず、近年の画像認識分野で目覚ましい成果を挙げている Convolutional Neural Network (CNN) を能動学習に適用するための新たな手法を提案し、その有効性を検証する。(ToDo: ここに手法の中身書くもの?)

### 1.3 本研究の構成

本論文の構成を以下に示す。

第1章で本研究の背景と目的を述べた。

第2章では、関連研究と本研究の位置付けについて述べる。

第3章では、本研究で提案する手法について述べる。

第4、5章では、提案した手法の有効性を確認するために行った実験の目的、設定、結果について述べる。

第6章では、本研究の結論、および展望について述べる。

### 1.4 本研究の貢献

(ToDo: ここもあやしい)

- 病理画像解析に能動学習を利用するためのシステムの構築

- 深層能動学習のために有効でかつ実用的なクエリ選択の戦略 Query-By-Dropout-Predictions を提案。

## 第 2 章

# 関連研究

### 2.1 画像認識における深層学習

#### 2.1.1 概観

深層学習とは、多層のニューラルネットワークを用いて高い識別性能を達成するモデルを構築する機械学習技術の一つである。Convolutional Neural Network (CNN) とは、画像認識分野において近年目覚ましい成果を挙げている多層ニューラルネットワークである。2012 年に大規模一般画像認識のコンペティションである ILSVRC で CNN を用いたチームが優勝して以来、画像に関する様々なタスクにおいて利用されている。従来の画像認識では、画像を認識するために有効だと考えられる特徴量を人手で設計し、その特徴量を用いて識別器を学習するというフレームワークであったのに対し CNN は、学習の過程で訓練データから識別に有効な特徴量を抽出する表現学習と、その特徴量の識別境界を決定する識別器の学習が同時に行われる点で特徴的である。

しかし一般に、CNN は非常に過学習に陥りやすく、大量のラベルつきデータセットを構築する必要があることが知られている。また、十分なデータ量がある場合でも、高い汎化性能を持つ CNN を学習させるにはいくつかの工夫を追加することが多い。本節の残りでは、過学習を防ぐために使用されるいくつかの手法について説明する。

#### 2.1.2 Dropout 正則化

一般に機械学習では、訓練データに対する過学習を防ぎ汎化性能を向上させるために、モデルの正則化が行われる。ニューラルネットワークの正則化手法の一つに Dropout[8, 9] がある。Dropout とは、学習の過程において、ニューラルネットワークの中間層のニューロンを一定確率でランダムに”drop”する、すなわち出力を 0 にする正則化手法である (Fig. 2.1.2)。これは、任意の  $n$  次元中間層の出力に対し、以下のようなマスクをかけることで行われる。



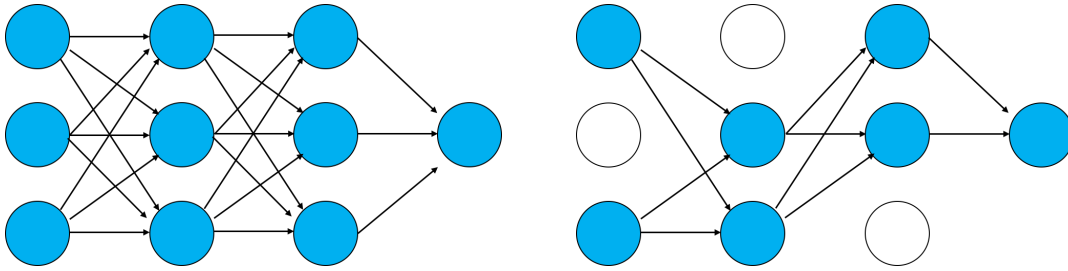


図 2.1. Dropout の模式図。(左) 通常のネットワーク (右) いくつかのニューロンを Drop されたネットワーク

(ToDo: 式)

$$y'_i = y_i \otimes M \quad (2.1)$$

$$M = [m_0, m_1, \dots, m_n], \quad m_i \sim \text{Bernoulli}(\pi)$$

これにより、あるニューロンが特定のニューロンにのみ過剰に依存してしまう共適合 (Co-adaptation) を防ぐことができる。また、指数的な数の部分ネットワークを同時に学習していると見なすこともでき、複数の部分ネットワークのアンサンブルを行っているのと同じ効果があるとも見なすことが出来る。

### 2.1.3 Data Augmentation

Data Augmentation とは、画像識別問題においてニューラルネットワークの過学習を防ぐために利用される正則化手法である。訓練時に使用する画像に対し、ラベル情報を保持する程度の変形 (主に幾何学的変形) を加えたものを入力して学習を行うことで、画像の様々な変形に対し不変な特徴量抽出、及び予測をすることを期待するものである。モデルに回転不変性を与えるために元画像を回転を加える、位置不変性を与えるために元画像に平行移動を加える等の変形がしばしば用いられる。

### 2.1.4 学習済みモデルからの転移学習

転移学習とは、ある問題を効率的に解くために、別の関連した問題のデータや学習結果を再利用する枠組みである。深層学習における転移学習とは、しばしば fine-tuning によるものを指すことが多い。fine-tuning とは、ニューラルネットワークを学習するため、別の問題もしくは別のデータセットで学習済みのモデルを初期値として利用して再学習を行うことである。近年では、大規模一般画像認識データセットである ImageNet によって学習済みの pre-trained model を fine-tuning させることで、幅広い画像認識タスクにおいて従来の手法と比較して高い性能を達成することが知られている。(ToDo: 引用) また、fine-tuning の場合、比較的少数の訓練データからでも優れた性能が得られることが示されている。(ToDo: 引用)

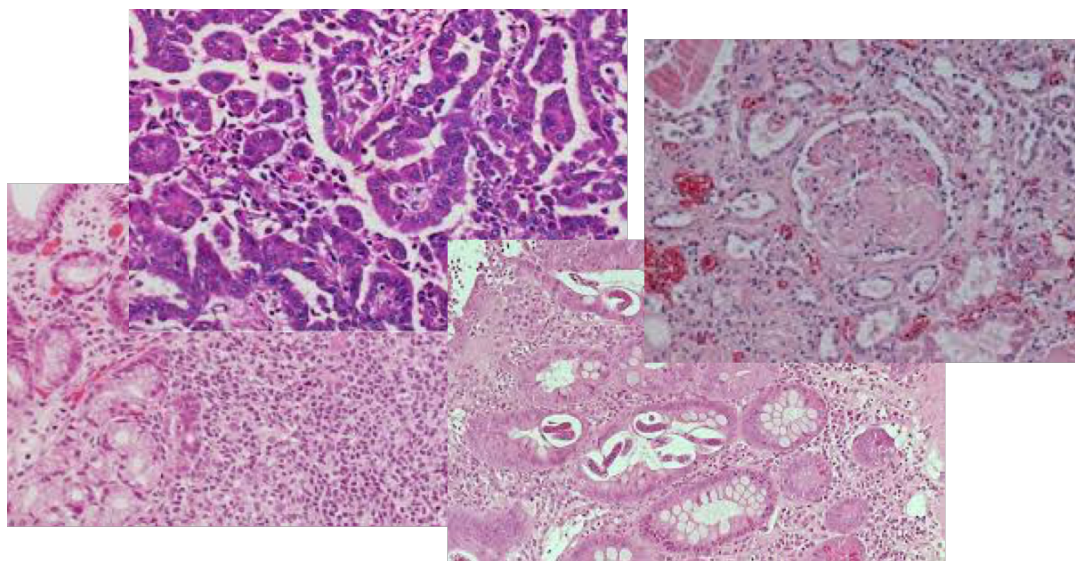


図 2.2. 病理画像の例。(ToDo: もうちょいいい感じのやつ)

## 2.2 病理画像解析

本節では、病理画像解析の性質と関連研究を述べる。

### 2.2.1 概観

第1章で述べたように、病理画像とは WSI と呼ばれる巨大なデジタル画像のことを指す。病理画像解析におけるアプリケーションは、大きく分けて3つに分類される [3]。自動診断による医師の補助、類似画像検索による医師の補助、画像と画像以外の情報（分子構造、遺伝子情報など）との関連性の解析、が挙げられる。中でも、近年の画像認識技術の向上により自動診断に関する研究が急速に増加している。(ToDo: 引用) 本研究でも、自動診断に位置づけられる癌（異常）の自動検知タスクに焦点を当てる。

### 2.2.2 WSI 全体の予測の出力方法

一般に、画像認識タスクにおいて用いられる画像はせいぜい  $1000 \times 1000$  以下である。数万ピクセル×数万ピクセルに及ぶ WSI 全てを一つのモデルに入力して一度に学習を行うのはデータ数、パラメータ数、どちらの面においても現実的ではない。また、癌などの異常部位というのは WSI のわずかな領域にのみ現れることも多く、解像度を落として使用しても期待する性能が得られない。そのため、多くの病理画像解析では、WSI を大量の画像パッチ ( $1000 \times 1000$  以下) に分割し、それぞれを個別に識別した結果を統合することで癌の有無を判定することが多い。個々の画像パッチを識別するモデルは、一般的な画像認識タスクで用いられるを使

用する。

### 2.2.3 使用される画像特徴量について

機械学習において特徴量抽出は重要な役割を持つ。一般に医療画像解析では、一般画像認識に使用される SIFT, SURF などの hand-crafted の画像特徴量が同様に利用されていた。(ToDo: 引用) また、病理画像はオブジェクトとしての特徴よりもテクスチャとしての性質を持つことから、HLAC, LBP などのテクスチャ解析における手法を採用する研究も盛んに行われていた (ToDo: 引用)。ただ、近年では一般画像認識における深層学習の成果を受けて、CNN を採用する研究が大半を占める。また、多くの医療画像解析ではラベル付きデータが少ないことから、ImageNet などのデータセットを学習済みの pretrained model の中間特徴量を特徴量として抽出し識別機は SVM などを利用する場合や、2.1.4 で述べたような転移学習を利用する場合がある。直観に反し、それらの中間特徴量は一般画像を認識するために学習されたものであっても、医療画像解析でも十分な性能を達成することができている。(ToDo: 引用)

一般に医療画像解析に共通している特徴として、獲得すべき不変性が多いということがある。(ToDo: 流れやや不自然) 病理画像では、各画像パッチは人手で切り取られたものであるため、向き、回転、位置に対しては不変であることが期待される。また、病理画像はそもそも染色液によって細胞を染色することで核などの病理学的特徴を見つけ出すための画像であるが、その染まり方は細胞毎、研究機関毎に大きく異なる (ToDo: 図があればいいかも) このことから、色相、輝度に対する不変性が強く必要とされる。(ToDo: 引用)

## 2.3 能動学習

この節では、能動学習についての基本的な概念および先行研究について述べる。また、近年の深層学習を識別器として利用した能動学習や、病理画像解析に能動学習を適用した例についても触れる。

### 2.3.1 基本的な概念

能動学習, Active Learning[?] とは、機械学習における一つの枠組みである。一般に、教師付き学習において識別精度の高いモデルを学習させるためには膨大なラベル付きデータが必要となる。そこで、能動学習では、与えられる大量のラベルなしデータから、モデルの更新に最も寄与する可能性のあるサンプルを選択し、それにアノテーションを付与することで、少ないアノテーションコストのもとでいかに高精度な学習モデルを作成するかを目的としている。データ自体は豊富にあるが、アノテーション付与のコストが高価であるような問題に用いられる。モデル更新に寄与する可能性があるかと判断する基準 (クエリ選考基準) は様々な側面から提案された手法があり、詳細を次項で述べる。

能動学習の基本的な流れを以下に示す

## 8 第2章 関連研究

1. モデル更新に寄与する可能性のあるサンプルを選択 (クエリ選択)
2. 人手, 専門家によってアノテーションを付与 (クエリ問い合わせ)
3. 付与されたアノテーションを利用し教師あり学習を実行 (再学習)

通常、学習が飽和するまでこれらのサイクルを繰り返す。

学習主体がクエリを問い合わせる問題設定は以下の3種類に大別される

### Stream-Based Selective Sampling

ストリーミングデータ (サンプルが次々に入力されていくような場合) に対するアプローチ。入力されるストリーミングデータに対してそれぞれのサンプルにラベルを付けるかを逐一判断し、モデル更新に寄与すると判断された場合 (大抵は閾値を設ける)、クエリとして問い合わせる方式である。

### Pool-Based Sampling

大量にラベルなしサンプルプールがまとまって与えられている問題に対するアプローチである。ストリーミングデータの場合とは違い、閾値を設けることなくプールの中から最もモデル更新に寄与するサンプルを選択することができる。また、サンプルデータの分布、およびデータ構造を考慮したクエリ選択が可能となる。

### Membership Query Synthesis

ストリーミングデータ、プールサンプルどちらの状況でも利用され、実際のデータを直接クエリとして利用するのではなく1つまたは複数のサンプルから新しい人口データを生成し人間に提示する事でラベル付けを行う方法である。

中でも、Pool-based Sampling の問題設定での研究が最も行われており、本研究でも大量のラベルなし病理画像データプールは予め与えられているものとする。

## 2.3.2 クエリ選考基準

本項では、モデル更新に寄与する可能性が高いと判断するためのクエリ選考基準として提案されているいくつかのものについて述べる。 $\mathcal{L}$  をラベル付きデータセット、 $\mathcal{U}$  をラベルなしデータセットとする。クエリはそれぞれの選考基準で使用されるスコアリング関数を最大にするサンプルを選択するものとする。

$$x^* = \operatorname{argmax}_{x \in \mathcal{U}} \operatorname{score}(x) \quad (2.2)$$

また、識別器のパラメータを  $\theta$  とし、サンプル  $x$  が与えられた時のクラスラベル  $y$  の予測確率を  $P_\theta(y|x)$  とする。

### Uncertainty Sampling

最もシンプルな戦略で、モデルにとって予測が最も曖昧であるようなサンプルをクエリとして選択する。データ  $x$  が与えられた時、ラベルの確率分布  $p(y|x)$  を出力する識別器であればよく、条件が緩いため幅広く利用されている。曖昧であることの定量評価として、事後確率

最大ラベルの確率が小さいサンプルを選択する Least Confident、事後確率最大ラベルの確率と、その次に大きなラベルの確率との差が小さいサンプルを選択する Margin Sampling、事後確率分布のエントロピーが大きいサンプルを選択する Entropy Sampling などが提案されている。多値分類の際はこれらはそれぞれ性質の異なるものになるが、二値分類の場合は全て等価な戦略となる。(ToDo: 式いい感じに)

$$score(x) = 1 - P_{\theta}(\hat{y}|x) \quad (\text{Least Confident}) \quad (2.3)$$

$$score(x) = P_{\theta}(\hat{y}_1|x) - P_{\theta}(\hat{y}_2|x) \quad (\text{Margin Sampling}) \quad (2.4)$$

$$score(x) = - \sum_i P_{\theta}(y_i|x) \log P_{\theta}(y|x) \quad (\text{Entropy Sampling}) \quad (2.5)$$

### Query By Committee

モデルのバージョン空間 (モデルの仮説空間内で現在のラベル付きデータに対して Consistent な領域) を縮小させるサンプルが更新に寄与するはずとする仮説に基づく戦略。実際にバージョン空間をすべて保持することは現実的には不可能であるため、近似的にこれを扱うための手法がいくつか提案されている。その中で代表的な手法が Query-By-Committee (QBC)[10] アルゴリズムである。バージョン空間を近似的に表現するために、現在のラベル付きデータ集合を使用して学習した複数のモデル (committee  $C = \{\theta^{(1)}, \dots, \theta^{(C)}\}$ ) を保持し、それぞれの committee の予測が最も不一致するサンプルを選択するという戦略である。不一致度の定量評価として、Vote Entropy、Average Kullback Divergence がある。Vote Entropy は、それぞれの committee がどのラベルだと予測 (投票) したかのばらつきを定量化したもので、次式で表される

$$score(x) = - \sum_i \frac{V(y_i)}{C} \log \frac{V(y_i)}{C} \quad (2.6)$$

$V(y_i)$  は committee の投票数、 $C$  は committee のサイズである。

Average Kullback Divergence は、それぞれ予測分布の確率分布の差異が平均的に大きいものを選ぶ手法である。

$$score(x) = - \frac{1}{C} \sum_{c=i}^C KL(P_{\theta^{(c)}} || P_C) \quad (2.7)$$

$P_C$  は committee 全体の平均の確率予測分布である。

### Expected Model Change / Expected Error Reduction

データにあるラベルがつけられた場合に、モデルが実際にどう更新されるかの期待値を求めることで、最善のサンプルを選択することができると考えられる。このような考えに基づき、可能性のあるラベルを全通り試すことでモデルの更新量が最大となるサンプルを選択するのが Expected Model Change と呼ばれる戦略で、また、モデルの更新量ではなくモデルの汎化誤



## 10 第2章 関連研究

差の減少量を評価することを目的とした戦略が Expected Model Change である。多くのケースでは、全サンプルのラベルを全通り仮定し、モデルの再学習を行う必要があるため膨大な計算コストを要する手法である。

### Variance Reduction

モデルの期待汎化誤差は以下のように分解することが出来る。

$$E_T[(\hat{y} - y)^2|x] = E_T[(y - E[y|x])^2] + (E_{\mathcal{L}}[\hat{y}] - E[y|x])^2 + E_{\mathcal{L}}[(\hat{y} - E_{\mathcal{L}}[\hat{y}])^2] \quad (2.8)$$

$E_{\mathcal{L}}[\cdot]$  はラベルセット  $\mathcal{L}$  に関する期待値、 $E[\cdot]$  は条件付き分布  $P(y|x)$  に関する期待値、 $E_T$  はそれぞれに関する期待値である。上記の右辺は、一項目から、ラベルノイズに関する項、モデルバイアスに関する項、モデルの分散に関する項を表している。これらのうち、学習によって変更できるのはモデルの分散のみである。そこで、モデルのパラメータの期待分散を小さくすることで、間接的にモデルの汎化誤差を小さくする戦略が Variance Reduction である。また、パラメータの分散はフィッシャー情報量  $I(\theta)$  の逆数によって下界を決定されるという統計的性質から、モデルのパラメータのフィッシャー情報量を最大化する (もしくは逆数を最小化する) サンプルを選択する問題に帰着される。パラメータが2つ以上ある場合、フィッシャー情報量はフィッシャー情報行列で表され、それらを扱うための計算量はパラメータ数に対して  $O(n^3)$  であるため、巨大なモデルに対しては、上記の Expected Model Change / Expected Error Reduction 程ではないが、計算量が膨大になる。

### Density Weighted Method

これまでの戦略とは異なり、単一のサンプルのみを評価するのではなく、周囲のサンプル、もしくは分布全体の構造を考慮する手法である。周囲にサンプルが多数ある場合と周囲にサンプルが存在しない場合を考慮すると、後者は外れ値のサンプルである可能性が高く前者のほうを選択する価値が高いと考えられる。そのようなヒューリスティクスを表現した手法の一つ Information Density がある。基本的に他の選択戦略と併せて使用される手法で、各定量評価に対して、類似度が高い他のサンプルがどれだけ存在するかを考慮した係数を掛け合わせることで重みをつける手法である。

$$score'(x) = score(x) \times \left( \frac{1}{U} \sum_{u=1}^U s(x, x_u) \right)^\beta \quad (2.9)$$

$s(\cdot)$  はサンプル間の類似度を算出する関数である。

### 2.3.3 その他の関連事項

(ToDo: 章立て微妙)

### サンプリングバイアス

能動学習全般に当てはまる問題として、学習主体が選択しアノテーションを付与されたデータ集合は、実際のデータ集合とは異なるという問題である。これをサンプリングバイアスと呼ぶ。これを緩和するために、しばしばクラスタリングなどにより実際のデータ集合に近づける手法が用いられる。

### バッチ型能動学習

多くの能動学習の研究では、クエリ問い合わせは一つのサンプルにのみ行われる。しかし、近年の機械学習アルゴリズムは計算コストが非常に大きいため、一度に複数のクエリ集合  $Q$  を問い合わせるバッチ能動学習に関する研究が増加しつつある。最もナイーブに考えたならば、2.3.2 で述べたような定量指標で上から  $Q$  個のサンプルを選ぶことになる。しかし、それらのサンプルには重複した情報が含まれる可能性が高い。brinker et al.、Xu et al. らはクラスタリング手法によって同一クラスター内からはクエリを選択しないという戦略を提案した [11]、[?]。また、Chen と Krause らは、クエリ選考基準を劣モジユラ関数によって表現することで、バッチ能動学習を劣モジユラ最適化問題に帰着することができると示した [12]。この時、貪欲法によって最適な組み合わせと比較した場合の  $1 - \frac{1}{e}$  近似が可能であることが保証される。しかし劣モジユラ関数の性質を担保するクエリ選考基準は限られており、線形識別器などのパラメータが少ないモデルにのみ適用可能な物が多い。

## 2.4 関連する先行研究と本研究の位置づけ

### 2.4.1 能動学習の病理分野への適用

能動学習を病理画像解析に利用した研究は複数存在する。複数人の WSI から切り出したパッチベースでの学習を行う病理画像解析では、クエリとして WSI 中の 1 領域を問い合わせるのが自然である。しかし、通常のデータ分布とは違い、独立同一分布 (i.i.d) でないという特徴がある。(ToDo: 回収しないと) NalisNik et al. は Uncertainty Sampling による能動学習を比較的単純な細胞核 segmentation タスクに応用した [13]。Doyle et al. は線形識別器を多数保持する query-by-committee によって癌検知の識別タスクに利用した [14]。また、Zhu et al. は画像パッチからテクスチャ特徴量を抽出し、線形識別器を採用して仮説空間の縮小をクエリ選考基準にすることで劣モジユラ最適化の枠組みを利用し、さらに k-means によるクラスタリングによって効率的にクエリを選択し、癌のステージの識別に利用した [15]。しかし、そのほとんどが人手によって設計された特徴量に対して線形の識別器を用いているものであるため、精度の点で十分であるとは言えない。この事実に関しては第 3 章の予備実験で検証する。

### 2.4.2 深層能動学習

能動学習の研究は、理論、実践的なアプリケーションどちらも線形識別器を用いているものが大半であるが、近年の深層学習の成果を受けて、深層学習と能動学習を組み合わせた研究も行われている [16, 17]。これらのアプローチは、Restricted Boltzmann Machines や stacked autoencoders などによって pre-training を行ったネットワークを用いて、Uncertainty Sampling によるクエリ選択を行うというシンプルなものである。(ToDo: もうちょいかく)

### 2.4.3 本研究の位置づけ

能動学習において深層学習を利用する研究の多くは、Uncertainty Sampling ベースのクエリ選考基準を用いている。これらのアプローチでは、多層ニューラルネットワークの予測の不確かさを定量的に評価するために、出力の確率分布のエントロピーが大ききなどが利用されることが多い。しかし、一般に、多層ニューラルネットワークは自らの予測に対する不確かさを陽に表すようにモデル化されていない。仮に予測が 0.9 という高い確率であったとしても、確信度が高いことを示すわけではないという問題がある。(ToDo: 非線形だからどこで間違っているかわからん。)(ToDo: ベイス CNN 触れる)

本研究では、不確かさを陽にモデル化しない多層ニューラルネットワークを能動学習に効率的に組み込むために手法を提案し、精度を犠牲にすることなく、安定してアノテーションコストを削減する病理画像解析システムを提案する。多層ニューラルネットワークにとってモデル更新に寄与すると考えられるサンプルを選択するために、Query-By-Committee のアイデアを基に Dropout と Data Augmentation を推論にも利用することにより近似的にそれらを達成するアプローチを取る。

## 2.5 まとめ

本章では、本研究に関連するいくつかの研究の枠組みと、先行研究について述べた。また、その中での本研究の位置づけを述べた。次の章では、具体的に本研究で提案するシステムを説明する。



## 第 3 章

# 提案手法

### 3.1 概要

本章では、本研究で提案するシステムについて説明する。本研究では、識別精度を犠牲にすることなくアノテーションコストを最小化する病理画像解析のための深層能動学習システムを構築する。高精度を達成するために識別器に Convolutional Neural Network を採用し、少ないラベル数でも学習を収束させるために ImageNet で学習済みの pretrained-network の重みから finetune して学習を行う。一般画像認識のために学習されたモデルでも、医療画像解析への転移学習は幅広いタスクに対して有効であると知られているため、妥当であると考えられる。

また、前章で述べた深層学習と能動学習の組み合わせの問題に対する解決するためのアプローチとして、Query-By-Dropout-Predictions を提案する。不確かさを定義しない多層ニューラルネットワークを効率的にパラメータを更新するためのサンプルを見つけるために、予測時にも Dropout を利用するアプローチである。さらに、医療画像解析において重要であると考えられる様々な変形に対する不変性の担保を考慮し、各 Committee の prediction 時にもランダムな Data Augmentation を利用することで、識別に有効なサンプルのみではなく不変性を確保するために有効であるサンプルをクエリとして選択することを考える。また、能動学習において一般に問題となるサンプリングバイアスを解決するためにクラスタリング手法を採用し、その際に用いる特徴量について妥当だと考えられるものを実験から選定した。

以下の節では、それぞれについて詳細を説明する。

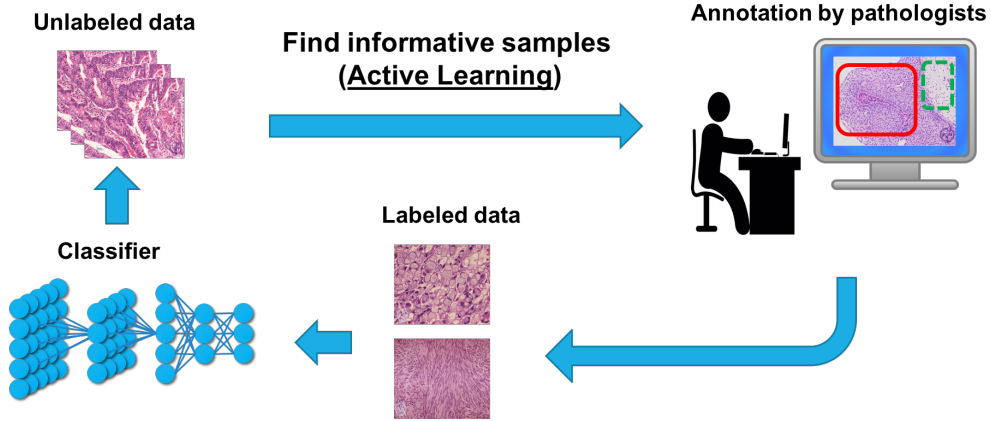


図 3.1. 本研究で提案するシステムの概念図

### 3.2 Query by Dropout predictions

多層ニューラルネットワークによって識別問題を解く際には、最終出力が各ラベルに対応する確率となる確率分布となるように設計して学習を行う。多層ニューラルネットワークを能動学習に利用する際にはこの確率分布のエントロピーの大きさなどを利用して不確かさを定量評価し、Uncertainty Sampling によってクエリ選択を行う。しかし、前章で述べたように、ニューラルネットワーク自体のパラメータに対する不確かさをモデルしているわけではない。このように不確かさを陽にモデル化しない識別器を利用した能動学習においては、バージョン空間を近似的に保持することで有効なサンプルを選択する Query By Committee を用いる事が多いが、多層ニューラルネットワークのように非常に計算コストの重いモデルを複数同時に保持し学習を行うことはメモリ、計算時間等様々な問題で現実的ではないという問題があった。そこで、本研究では、多くの深層学習のアーキテクチャで正則化の目的で利用される Dropout によって、本体のネットワークからサンプリングされた部分ネットワークを Committee とみなし、それらの予測の不一致度を利用して近似的に Query-By-Committee を行う手法を提案する (Fig.3.2)。つまり、本体のネットワーク  $\mathcal{M}$  から生成された部分ネットワーク  $\mathcal{C} = \{\mathcal{M}_{p_1}, \mathcal{M}_{p_2}, \dots, \mathcal{M}_{p_c}\}$  を用いて以下の Average Kullback Divergence を計算する。

$$score(x) = -\frac{1}{C} \sum_{c=i}^C KL(P_{\theta(c)} || P_C) \quad (3.1)$$

(ToDo: disagreement だけでなく不確かさも利用する話どこでしょうかな) これらの部分ネットワークが Committee として利用されるためには、それらが訓練データに対しては Consistent であり、かつ、それぞれの未知データに対する出力には分散を持つ必要がある。この性質について実験で検証し、Uncertainty Sampling のみを利用する方法と比較する。

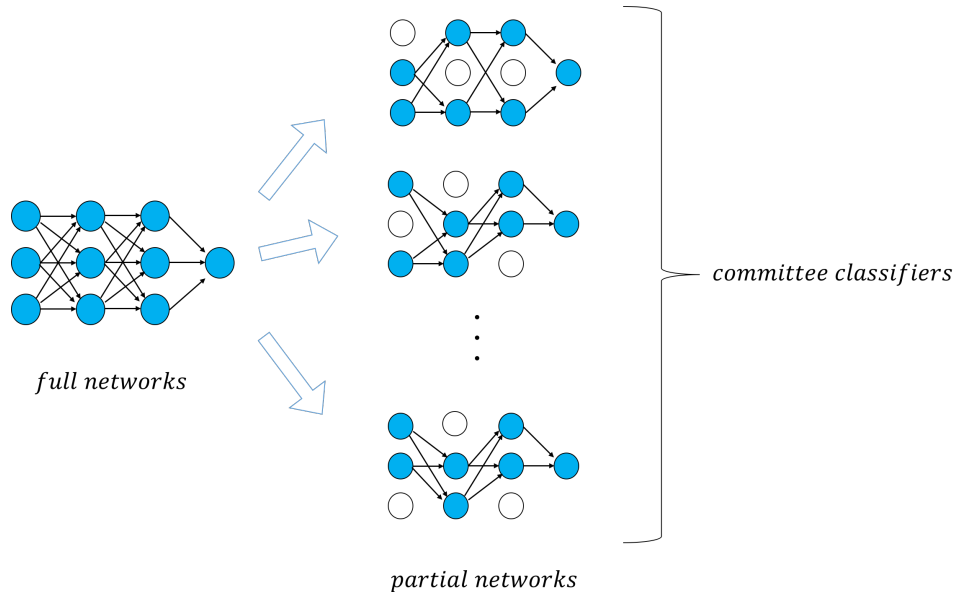


図 3.2. Dropout によってサンプリングされた部分ネットワークによって Committee を形成する。

### 3.3 推論時での Data Augmentation の利用

CNN を学習させる際、訓練時に画像に対して Data Augmentation を利用することで不変性を獲得させ、結果的に汎化性能を向上させる工夫が用いられる。医療画像解析では、一般画像認識と比較してモデルが獲得すべき不変性が多く存在する。回転不変性、色相不変性、などがある。ラベル付きデータが少ない場合には特に有効であると知られており、本研究でも識別器の学習に使用する。上記で Dropout によって部分ネットワーク Committee を生成することで不一致度を図る手法を提案したが、本研究ではそれだけではなく、Data Augmentation を推論時にも利用することで、

### 3.4 提案手法の動作原理

pre-trained network を利用する。基本的には dropout は fully connected 層で行うよってそれらは識別境界を大きく変更させるためのサンプルを取得することができる。Data augmentation を利用することで

### 3.5 バッチ型能動学習への拡張

能動学習では、クエリ問い合わせを行いラベルデータを拡充するごとに、識別器を from scratch から学習させるのが一般的である。深層学習器を能動学習に利用する場合、一度の再

学習に大きな計算コストがかかってしまうため、各クエリ問い合わせにおいて複数のサンプルにラベルを付与するバッチ型能動学習を採用するのが適当であると考えられる。一度に投げるクエリ内での情報の重複を避けるために何らかの工夫をする必要があるが、2章で述べたような劣モジュール関数を設計するのは深層学習を利用する場合計算コストの観点から難しい。そこで、本研究ではクラスタリングにより同一のクラスターに属するサンプルは2つ以上選択しないとすることで、情報の重複を避ける。また、クラスタリングに使用する特徴量の候補として、以下の3つを考慮した。

#### hand-crafted feature

2章で述べたように、病理画像解析にはテクスチャ解析で用いられる技術がしばしば使用される。ここでは、パターンベースの特徴量で、位置不変性と輝度変化への頑健性を有する Local Binary Pattern (LBP) を採用する。実際には、回転不変性を担保させるように LBP を改良した improved LBP を利用した。

#### CNN の中間特徴量

Imagenet で学習された特徴量は様々なタスクに有用であるとされており、多数の医療画像解析で利用されている。本研究では GoogleNet の中間特徴量 512 次元を利用した。

#### compact bilinear pooling による特徴量

近年の深層学習を用いたテクスチャ解析の研究で、Bilinear Pooling によって空間情報をなくした特徴量がしばしば利用される。これは、CNN の中間特徴量の相関行列を計算し空間方向に平均を取ったものである。

$$G_{ij} = \sum_k F_{ik} F_{jk} \quad (3.2)$$

また、一般に CNN の特徴量次元（チャンネル数）は 256 512 の大きな値であるため、そのまま計算した場合非常に高次元な値になってしまう。そこで、その近似手法である Compact Bilinear Pooling[18] を本研究では CNN を用いたテクスチャ特徴量として採用する。

これらを比較するために、それぞれの特徴量を用いて病理画像データセットを K-means によってクラスタリングを行った際の各クラスター内のラベルの不純度を平均を計算した。不純度が小さいほどクラスター内でのばらつきが小さく良い特徴量だと言える (ToDo: ちょいあやしいか)。データセットは 100000 枚の病理画像からなり、癌と正常の割合は均等に調整した。使用したデータセットの詳細は第5章で説明する。表 3.5 に示すように、CNN を用いたテクスチャ特徴量である Compact Bilinear Pooling をクラスタリングに用いるのが妥当であると考えられる。

### 3.6 アルゴリズムの詳細

Algorithm.1 に本研究で提案するアルゴリズムの詳細を示した。

表 3.1. 比較実験の結果

手法	Accuracy
LBP	0.396
CNN の中間特徴量	0.335
Compact Bilinear Pooling	<b>0.330</b>

**Algorithm 1** Deep Active Learning for Pathological Image Analysis**Input:** unlabeled dataset  $\mathcal{U}$ , clustering size  $k$ , active sampling size  $K$ ,**Output:** paramters of network  $\mathcal{M}$ 

- 1: Initialize pre-trained network  $\mathcal{M}_{pretrained}$
- 2: Perform k-means clustering with texture representation  
and devide  $\mathcal{U}$  to disjoint clusters  $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_k$
- 3: Initialize labeled dataset  $\mathcal{L}$  from each disjoint clusters
- 4:  $\mathcal{M} \leftarrow \text{train}(\mathcal{M}_{pretrained}, \mathcal{L})$
- 5: **repeat**
- 6:   sample committees  $\mathcal{C} = \{\mathcal{M}_{p_1}, \mathcal{M}_{p_2}, \dots, \mathcal{M}_{p_c}\}$  from  $\mathcal{M}$  using *dropout*
- 7:   **for** each  $x_i \in \mathcal{U}$  **do**
- 8:      $\text{score}(x) = \text{disagreement\_score}(x, \mathcal{C})$
- 9:   **end for**
- 10:    $\mathcal{Q} \leftarrow \emptyset, \mathcal{D} \leftarrow \emptyset$
- 11:   **while**  $\text{len}(\mathcal{Q}) < K$  **do**
- 12:      $x^* = \text{argmax}_{x \in \mathcal{U}} \text{score}(x)$
- 13:      $\text{idx} = \text{cluster}(x^*)$
- 14:     **if**  $\text{idx} \notin \mathcal{D}$  **then**
- 15:        $\mathcal{Q} \leftarrow \mathcal{Q} \cup \{x^*\}, \mathcal{D} \leftarrow \mathcal{D} \cup \{\text{idx}\}$
- 16:     **end if**
- 17:   **end while**
- 18:   Query labels for  $\mathcal{Q}$
- 19:    $\mathcal{L} \leftarrow \mathcal{L} \cup \mathcal{Q}$
- 20:    $\mathcal{M} \leftarrow \text{train}(\mathcal{M}_{pretrained}, \mathcal{L})$
- 21: **until** perfomance is satisfactory

## 第 4 章

# 実験 1 : MNIST を用いた予備実験

### 4.1 概要

本章では、本研究で提案する Query-By-Dropout-Predictions の有効性を検証するために MNIST に対して行った実験について説明する。Dropout によってサンプリングされた各部分ネットワークの出力が未知データに対して分散を持つのかを検証し、能動学習のクエリ選考基準として有効であるかを確認する。

### 4.2 実験設定

#### 4.2.1 データセットについて

MNIST は  $28 \times 28$  ピクセルの手書き数字のデータセットである。7 万枚の画像からなり、そのうちの 6 万枚は訓練画像、残りの 1 万枚はテスト画像として利用される。それぞれの画像は 0 9 までの数字ラベルが割り当てられているが、能動学習の状況を再現するため、クエリとして問い合わせられるまではラベルへのアクセスが与えられない状況を設定する。

#### 4.2.2 実験の詳細

表 4.1. MNIST の実験に使用した CNN の構造を示す。

layer	size	activation
Convolution	$32 \times 4 \times 4$	relu
Convolution	$32 \times 4 \times 4$	relu
Max Pooling		
dropout		
fully connected	128	relu
dropout		
fully connected	128	relu

識別器には CNN を利用する。その構造を以下に示す 4.2.2(Keras の example プログラムを参考にした)。比較的小さな CNN であるため、クエリ問い合わせによってラベルが追加されるごとに from scratch からの再学習を行うことにする。多層ニューラルネットワークの中では小さいモデルではあるものの、ラベルを一つずつ追加して再学習を行うのは計算コストが大きいため、バッチ型能動学習を採用する。ここでも、同一クエリ内での情報の重複を避けるためクラスタリングを行う。MNIST は 784 次元の画像で比較的小さいため、元の画像情報をそのまま特徴量として k-means によるクラスタリングを行う。committee サイズは 10、k-means のクラスター数  $K$  は 100、一度に選択するクエリ  $Q$  のサイズは 10、再学習の epoch 数は 100 に設定した。また、クラスターの代表サンプルから無作為にサンプリングされた 20 個のサンプルをラベルを付与して学習初期のラベルつきデータとして実験を開始した。訓練時には crop 領域をランダムにずらす Augmentation を利用した。各クエリ問い合わせ毎に 10000 枚のテストデータに対する識別精度を計測し、ラベルを付与されたデータが 1000 に到達するまで実験を続けた。実験ごとのばらつきを考慮し、同一の実験を 3 回行いその平均と標準偏差を計算した。

### 4.2.3 比較手法について

提案する Query-By-Dropout-Predictions の性能を比較するため、いくつかの手法について実験を行った。クエリの選考基準以外は全ての設定を揃えて実験を行った。

#### Random Sampling

この実験のベースラインと言える選択基準。各クエリ問い合わせ毎にランダムにサンプルを選択してラベル付きデータセットに追加する。

#### Uncertainty Sampling

推論時に Dropout を使用せずに単一の予測分布のエントロピーを利用する。

$$score(x) = - \sum_i P_\theta(y_i|x) \log P_\theta(y|x) \quad (4.1)$$

#### Query-By-Dropout-Predictions + Uncertain Sampling

推論時に Dropout を利用し、複数の prediction を出力しそれらの平均の不確かさと不一致度を基準として利用する。不確かさには予測分布のエントロピー (Entropy Sampling) を利用する。

$$score(x) = - \frac{1}{C} \sum_{c=1}^C KL(P_{\theta^{(c)}} || P_C) - \sum_i P_C(y_i|x) \log P_C(y|x) \quad (4.2)$$

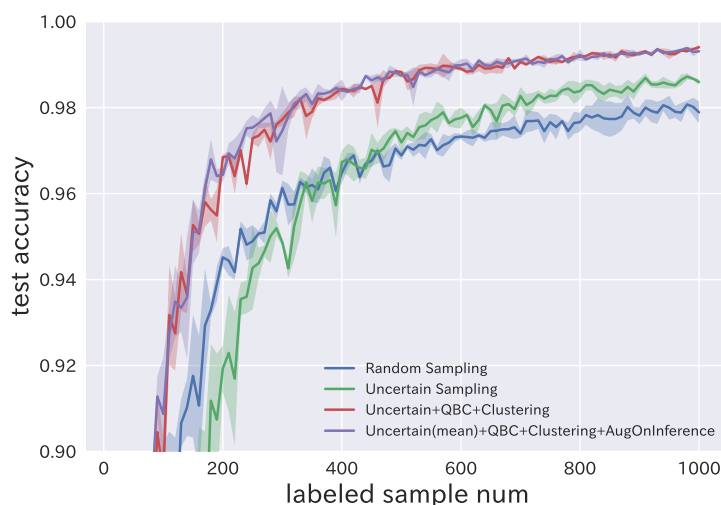


図 4.1. 各手法を利用した場合のラベル付きサンプル数の増加に対するテスト精度の変化を示した図

#### Query-By-Dropout-Predictions + Uncertain Sampling + 推論時 Data Augmentation

上記のクエリ基準に加え、推論時にも Data Augmentation を利用することで特徴抽出層の学習に有効だと考えられるサンプルも選択できるようにする。

### 4.3 実験結果

本項では上記で述べた実験の結果を示す。クエリ選考基準として提案手法、比較手法それぞれを使用した際の、ラベル付きサンプル数の増加に対するテスト精度の変化のグラフを図 5.3 に示す。また、テストデータの識別精度が 90%、95%、98% を超えるのに要したラベルの数を表 4.3 に示す。提案した Query-By-Dropout-Committee の手法が Uncertain Sampling のみを使用した場合と比較して性能が良いことがわかる。これは上記に述べた通り多層ニューラルネットワークが正しくモデルの不確かさを表現できていないからだと推測される。さらに、推論時にも Data Augmentation を利用することで、さらに少ないラベルで高精度を達成していることがわかる。

### 4.4 考察

以下にラベル付きデータセットサイズが 100 の時にクエリとして選択されたサンプルを示す。図?? QBC が Deep と相性が良いことがわかった。また、それを示すために訓練データとテストデータに対して Dropout を利用した場合のそれぞれの分布の分散を比較した。



表 4.2. それぞれのクエリ選考基準を利用した場合にテスト精度を達成するために要したラベルの数

	90%	95%	98%
Random	110 $\pm$ 20	240 $\pm$ 10	900 $\pm$ 70
Uncertain Sampling	180 $\pm$ 5	270 $\pm$ 20	600 $\pm$ 50
Uncertain + QBC + Clustering	100 $\pm$ 10	150 $\pm$ 20	320 $\pm$ 10
Uncertain + QBC + Clustering + DA	90 $\pm$ 0	150 $\pm$ 5	300 $\pm$ 30

## 第 5 章

# 実験 2：病理画像データセットを用いた実験

### 5.1 概要

MNIST の実験から、提案したクエリ選考基準は、既存手法よりも CNN にとって識別率向上に寄与するサンプルを選択できていることが確認できた。本章では、本研究で提案するシステムを用いて実際に病理画像データセットに対して行った実験について説明する。

### 5.2 実験設定

#### 5.2.1 データセットについて

本実験では、Camelyon Grand Challenge[7] にて公開された Camelyon データセットを利用した。Camelyon データセットは 1000 枚の WSI からなり、乳癌のリンパ節転移を自動で検出する識別器を生成することを目的としたデータセットである (ToDo: 図)。各 WSI は一枚あたり  $100,000 \times 100,000$  ピクセルの大きさで、細胞組織を含む画像パッチは約 10,000 から 400,000 枚になる。また、それら全ての画像パッチに対して癌か正常かの二値のラベルが付与されている。ここでも、MNIST で行った実験と同様に各画像パッチのラベルは、クエリとして問い合わせない限り与えられない状況とする。

#### 5.2.2 実験の詳細

Mnist の実験の際とは異なり、訓練用 WSI に含まれる全ての画像パッチをラベルなしデータセット  $\mathcal{U}$  として利用した場合、その数は約 6,000,000 サンプルに登ってしまう。このサンプルプールから、あるクエリ選考基準を最大化するサンプルを選択するのは計算コストの点から実用上困難である。そこで、クエリ問い合わせ毎に利用可能な画像パッチ全体から一部をサンプリングすることで  $\mathcal{U}_i$  を作成し、この中からモデル更新に寄与するサンプルを選択するようにする。各  $\mathcal{U}_i$  のサイズは 50,000 に設定した。

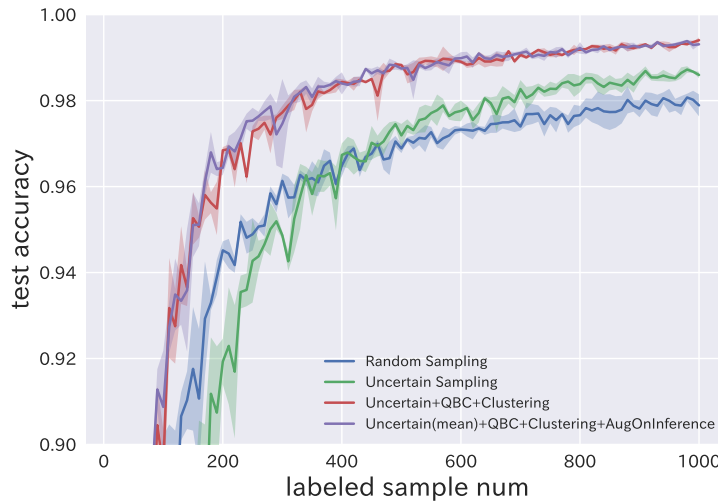


図 5.1. 各手法を利用した場合のラベル付きサンプル数の増加に対するテスト精度の変化を示した図

識別機に用いる CNN は GoogLeNet を採用した。他の一般画像認識で利用される CNN のアーキテクチャの中でも比較的計量で、医療画像解析ではしばしば用いられるモデルであることから選択した。(ToDo: 引用と図) 少ないラベルで高精度を達成するために、ImageNet で学習済みの pretrained-model を初期値として、転移学習によって学習を行う。

本実験でも、同一クエリ内での情報の重複を避けるためクラスタリングを行う。第 3 章の予備実験から、クラスタリングに使用する特徴量は Compact Bilinear Pooling を用いたテクスチャ特徴量が有効だと推測される。committee サイズは 10, k-means のクラスター数  $K$  は 1000、一度に選択するクエリ  $Q$  のサイズは 100 に設定した。Mnist での実験のように各 iteration での学習回数を固定にせず、より現実的な状況に近づけるため、与えられた比較的小さいバリデーションセットの性能が変化しなくなるまで学習を行う、という設定にした。

また、クラスターの代表サンプルから無作為にサンプリングされた 100 個のサンプルにラベルを付与して学習初期のラベルつきデータとして実験を開始した。各クエリ問い合わせ毎に 10000 枚のテストデータに対する識別精度を計測し、ラベルを付与されたデータが 10000 に到達するまで実験を続けた。実験ごとのばらつきを考慮し、同一の実験を 3 回行いその平均と標準偏差を計算した。

## 5.3 実験結果

## 5.4 考察

## 第 6 章

# 結論

本章では、本論文の結論および課題と今後の展望について述べる。

### 6.1 結論

本研究では、病理画像解析において問題となっている膨大なアノテーションコストを緩和するためのアプローチとして、能動学習を取り入れることを提案した。また、近年の画像認識で成果を挙げている深層学習を、能動学習における識別機として利用する際のクエリ選考基準を考案した。モデルのバージョン空間を縮小させるサンプルを選択する Query-By-Committee を Dropout によって近似的に再現し、それらの不一致度を定量化することで CNN のパラメータ更新に寄与するサンプルを効率的に探索が可能であることを、簡易的なデータセットと大規模病理画像データセットでの実験で検証した。さらに、汎化性能を向上させるために通常は訓練時にのみ使用される Data Augmentation を推論時にも使用することで、特徴抽出のために有効であると考えられるサンプルを取得する手法を提案しその病理画像における顕著な性能を示した。

本研究で得られた知見は以下のとおりである。

### 6.2 今後の課題

#### 6.2.1 ラベルなしデータの活用

本研究では、モデルの分散を最小化するラベル付きデータセットを作成するという目的で実験を行っており、作成したあとに自由にハイパーパラメータを設定して性能を調整するという方針を取った。しかし、半教師付き学習のようにラベルなしデータも学習に利用することでさらに効率的にサンプル選択をすることができる可能性がある。

### 6.2.2 継続的な finetune による学習

本研究では、モデルが各再学習において過学習してしまうことを恐れ、ラベル付データが追加された後にモデルを pre-trained のパラメータに初期化していたが、本来はその直前に追加されたサンプルはその時点でのモデルパラメータにとって最適であったものなので、初期化してしまった場合してしまった場合真のサンプルとずれてしまう可能性がある。また、再学習にはその分時間もかかるため、継続的な finetune をするための工夫を考える余地は残されている。

## 参考文献

- [1] Liron Pantanowitz. Digital images and the future of digital pathology. *Journal of pathology informatics*, Vol. 1, , 2010.
- [2] Metin N Gurcan, Laura E Boucheron, Ali Can, Anant Madabhushi, Nasir M Rajpoot, and Bulent Yener. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, Vol. 2, pp. 147–171, 2009.
- [3] Daisuke Komura and Shumpei Ishikawa. Machine learning methods for histopathological image analysis. *CoRR*, Vol. abs/1709.00786, , 2017.
- [4] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *arXiv preprint arXiv:1702.05747*, 2017.
- [5] Scott Doyle, Shannon Agner, Anant Madabhushi, Michael Feldman, and John Tomaszewski. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pp. 496–499. IEEE, 2008.
- [6] M Murat Dundar, Sunil Badve, Gokhan Bilgin, Vikas Raykar, Rohit Jain, Olcay Sertel, and Metin N Gurcan. Computerized classification of intraductal breast lesions using histopathological images. *IEEE Transactions on Biomedical Engineering*, Vol. 58, No. 7, pp. 1977–1984, 2011.
- [7] Oscar Geessink, Péter Bándi, Geert Litjens, and Jeroen van der Laak. Camelyon17: Grand challenge on cancer metastasis detection and classification in lymph nodes, 2017.
- [8] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.
- [9] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*, Vol. 15, No. 1, pp. 1929–1958, 2014.

- [10] H Sebastian Seung, Manfred Opper, and Haim Sompolinsky. Query by committee. In *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 287–294. ACM, 1992.
- [11] Klaus Brinker. Incorporating diversity in active learning with support vector machines. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 59–66, 2003.
- [12] Yuxin Chen and Andreas Krause. Near-optimal batch mode active learning and adaptive submodular optimization. In *ICML (1)*, pp. 160–168, 2013.
- [13] Michael Nalisnik, Mohamed Amgad, Sanghoon Lee, Sameer H Halani, Jose Velazquez Vega, Daniel J Brat, David A Gutman, and Lee AD Cooper. Interactive phenotyping of large-scale histology imaging data with histomicsml. *bioRxiv*, p. 140236, 2017.
- [14] Scott Doyle, James Monaco, Michael Feldman, John Tomaszewski, and Anant Madabhushi. An active learning based classification strategy for the minority class problem: application to histopathology annotation. *BMC bioinformatics*, Vol. 12, No. 1, p. 424, 2011.
- [15] Yan Zhu, Shaoting Zhang, Wei Liu, and Dimitris N Metaxas. Scalable histopathological image analysis via active learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 369–376. Springer, 2014.
- [16] D. Wang and Y. Shang. A new active labeling method for deep learning. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pp. 112–119, July 2014.
- [17] Jiming Li. Active learning for hyperspectral image classification with a stacked autoencoders based neural network. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pp. 1062–1065. IEEE, 2016.
- [18] Yang Gao, Oscar Beijbom, Ning Zhang, and Trevor Darrell. Compact bilinear pooling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 317–326, 2016.





# 謝辞

本研究は、東京大学大学院 情報理工学系研究科 創造情報学専攻の原田達也教授のご指導のもとに行われました。原田達也教授、牛久祥孝講師をはじめとして、原田・牛久研究室のメンバーの皆様から多大な支援を受けて本論文を完成させることができました。

原田達也教授の研究に対する意識の高さ、視野の広さは研究を行う上で大きな刺激となりました。非常にお忙しいにも関わらず、的確な指導をしていただき、原田教授のご指摘を通して自分の見逃していた部分に気づかされることも多く、とても勉強になりました。心より感謝しております。また是非テニスしましょう。

牛久祥孝講師は生徒に対する面倒見が良く、研究会での幅広い知識をバックグラウンドにしたアドバイスも的確を射ていたのが印象的です。ベンチャー立ち上げは流れてしまいましたが、もし自分が会社疲れに疲れたらまた相談に乗っていただけたら嬉しいです。

中村衛助教には、研究室内の事務作業を担当していただき、感謝しております。何度も勤務表の手続きでご迷惑をおかけして申し訳ありませんでした。

特任研究員の Tejero de Pablos Antonio さん、黒瀬優介さんの研究会での鋭い指摘にはいつもハッとさせられました。技術補佐員の金子葵さん、加治佐美由希さんは我々に快適な研究環境を準備すべく多面にわたり尽力していただき、感謝しております。

大学院博士課程の森友亮さん、日高雅俊さん、棕田悠介さん、Huang KaiKai さん、南里卓也さん、兼平篤志さん、葉 浩さん、加藤大晴さん、Tang Yujin さん、Li Yang さん、金子卓弘さん、Mohammad Reza MOTALLEBI さん、温 穎怡さんには熱心に研究に取り組む姿から研究生活に関する多くのことを学ばせていただきました。また、研究の進め方の手本を見せていただいたばかりでなく、研究室での充実した生活を提供してくださったことに心より感謝いたします。棕田さんには機械学習手法に関するお話や、新しい論文に関する議論をしていただき大変勉強になりました。ありがとうございました。日高さんには自分が修士に入りたての頃に alsarc 2016 を通して初歩的なことを色々教えていただきました。また、それ以降もサーバや実装に関する質問を何度もさせていただきました。ありがとうございました。加藤さんは、GAN などの楽しい論文について共有していただきお話するのが楽しかったです。また、もう一つの顔であるぱろすけさんは、自分がプログラミングを始めるきっかけになった方でもあるのでそちらに関してもとても感謝しています。

修士課程二年目の同期である井関茜さん、齋藤邦章くん、床爪佑司くん、山本将平くん、早川顕生くん、木倉悠一郎くん、高田一真くん、YanTennin くん、石原弘之さん、渡辺康平くん、張仁彦くん、Hanna Tseran さんは、共に卒業を目指し研究に取り組む同輩として心強く感じ

### 30 参考文献

させていただきました。研究室では楽しい会話を提供してくださり、心の支えになりました。

修士課程一年目の後輩である岡田英樹くん、高柳臣克くん、唐澤拓己くん、金山哲平くん、航くん、町田龍昭くん、荒瀬晃介くん、新井棟大くん、稲田修也くん、James Borg くん、張徳軒くんは、普段の話相手になっていただけるだけでなく、それぞれが違う興味をもっており良いインスピレーションを受けました。

学部4年の石川輝くん、上原康平くん、宇佐美峻くん、河合里咲さん、田中幹大くん、津野蒼くん、野口敦裕くん、松浦寿彦くんは、自分がB4の頃と比べられないくらい優秀で、全員が真摯に研究に取り組んでおり、自分も負けていけないなという刺激を受けました。

本論文の完成は、研究室の皆様のご協力がなければ到底なしえないものでした。最後に、改めて皆様に感謝の意を表すとともに、どんな時にもあたたかく励まし支え続けてくれた家族に感謝し、本論文の謝辞とさせていただきます。