

# Author Guidelines for ECCV Submission

Anonymous ECCV submission

Paper ID \*\*\*

**Abstract.** Along with the spread of digital pathology, many types of research have been conducted to support the diagnosis of a pathologist using computational image analysis. Generally, it is often necessary to construct large-scale labeled dataset to acquire a model that achieves high classification accuracy. However, the cost of preparing such dataset for pathological images is quite expensive. To make annotations to medical images such as pathological images requires highly specialized knowledge and all the annotation work must be performed by a doctor.

Therefore, in this research, we aim to construct a system that acquires high accuracy model for pathological images while suppressing annotation cost. For this approach, we adopt an Active Learning framework, which automatically select the informative samples from an unlabeled image set and request people to annotate. In addition, as a new method for applying Convolutional Neural Network (CNN) to active learning, we propose Query-By-Dropout-Predictions (QBDP) that utilizes the disagreement degree of prediction of partial network sampled by Dropout. As a result of the experiment, in general image dataset, we show that our method outperforms conventional method. Furthermore, we applied proposed system to the large-scale pathological image dataset and demonstrate its effectiveness.

...

**Keywords:** We would like to encourage you to list your keywords within the abstract section

## 1 Introduction

### 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 スライド当たりの画素数が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 [2–4]。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、機械学習手法を用いて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。これは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 Deep Convolutional Neural Network (DCNN) などの深層学習では特に顕著である。

病理画像などの医療画像に対するラベルの付与には高度な専門的知識を要するため、全てのアノテーション作業を医師が行わなければならず、学習データセットの作成コストが非常に高い。また、特に病理画像は WSI 1 枚あたりの画素数がギガピクセルに及ぶほど巨大であるため、詳細なアノテーションの付与には長時間に及ぶ集中力が求められる。例えば、Camelyon Grand Challenge [7] にて公開された Camelyon Dataset では、1,000 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。これらのアノテーションは WSI 1 枚につき熟練の病理医が 1 時間かけて行ったとされ、延べ 1,000 時間をこのデータセット作成に費やしたという報告がある。

学習データセットの作成コストが問題となるのは医療画像解析に限らず、工場での異常検知や、詳細画像認識など様々な状況で起こる。機械学習では、少量のデータから効率的に学習するための枠組みが存在する。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 [8] 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。またこの例えを用いると、教師なし学習は問題集のみが与えられ解答が一切与えられない状況だと言える。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、学習の過程で教わりたい問題の答えだけを教師に聞きながら学習を行う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付与してもらい、そのデータを利用して再学習を行うというサイクルを繰り返すことでアノテーションコストを抑え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年々増加しており、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。

しかし能動学習の研究は、一部の例外を除き高次元のデータへのスケーラビリティに欠く手法が多い。病理画像では、WSI 1 枚あたりに大量の画像パッチを含むうえに、それらの画像パッチ 1 枚 1 枚も高次元な画像データであるため、扱いが非常に難しいと言える。また、識別器には線形識別器を仮定している研究が多く、計算コストの面から CNN のように膨大なパラメータ数を持つモデルを利用できない場合が多いことも問題として挙げられる。

## 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に対して大きな負担となるアノテーションコストを抑えつつモデルの識別精度を担保するために、学習の過程で識別器とアノテーターとのインタラクションを通じてラベル付き学習データを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。また、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として，Dropout によってサンプリングされた部分ネットワークの予測の不一致度を利用する Query-By-Dropout-Predictions (QBDP) を提案し，その有効性を検証する。

### 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

- 大規模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
- 深層能動学習において，現実的な計算コストで効率的にクエリを選択するアルゴリズム Query-By-Dropout-Predictions の提案
- 不一致度を定量化するための推論時に，Dropout だけでなく Data Augmentation も併用することでさらに効率的なサンプル選択を可能にする手法の提案
- 大量の画像パッチからなる病理画像データセットを能動学習で扱うために，ラベルなしデータセットをクエリ選択のたびにサンプリングすることで膨大なデータセットに制限を加えることなく現実的な時間でクエリ選択を行う手法の提案

## 2 Related Work

本研究では，深層能動学習によってアノテーションコストを抑えつつ大規模病理画像解析を行うためのシステムを構築する。本章では，それらに関連する研究の枠組みについて述べる。2.1 節では画像認識で目覚ましい成果を挙げている深層学習において，汎化性能を向上させるためのいくつかの工夫について述べる。2.2 節では病理画像解析の性質と先行研究について説明する。2.3 節では能動学習に関する基本的な概念について説明する。2.4 節では本研究に関連する先行研究を概観し，本研究の位置付けを述べる。

### 2.1 画像認識における深層学習

**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 1 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 [9] 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

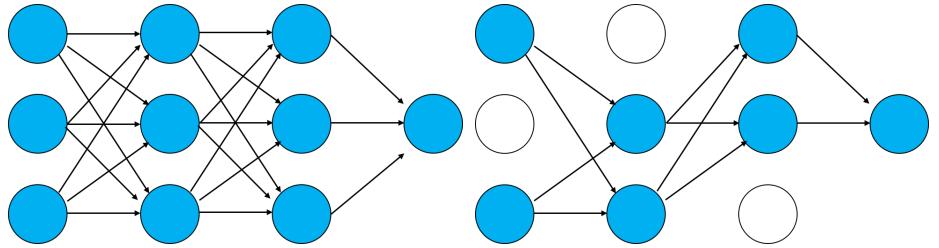


Fig. 1: Dropout の模式図. (左) 通常のネットワーク (右) いくつかのニューロンが Drop されたネットワーク

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, それらの工夫の中でも本研究に関連するものについて説明する.

**Dropout** 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 1 つに Dropout [10, 11] 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 (図. 1). これは, 任意の  $n$  次元中間層の出力  $y_i$  に対し, 以下のようなマスク  $M$  との要素積  $\otimes$  を取ることで行われる.

$$y'_i = y_i \otimes M$$

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

$y'_i$  は次の層への入力として使用される.  $m_i$  はそれぞれ, パラメータ  $\pi$  のベルヌイ分布から順伝播毎にサンプリングされる値である.

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, 指数的組み合わせの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.

**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汎化性能を向上させるために利用される手法である. 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变な特徴量を抽出して予測することを期待するものである. モデルに回転不变性を与える

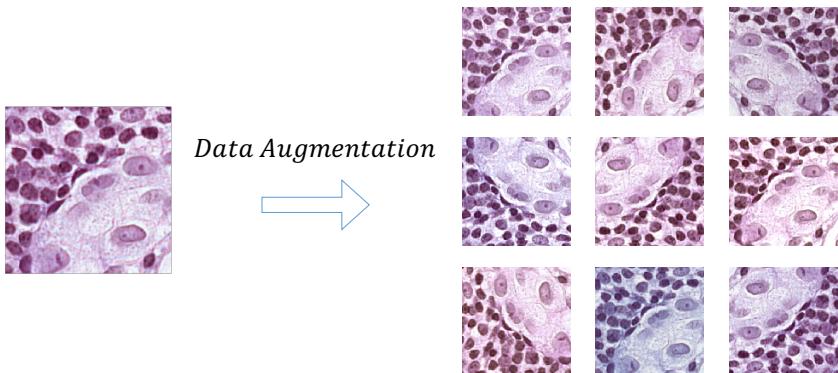


Fig. 2: 1枚の病理画像パッチに対して Data Augmentation を行った例。ここでは、Random Crop, Random Rotation, Random Flip, Random Color Augmentation を使用した。

ために元画像に回転を加える、位置不变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。例を図 2 に示す。

**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に別のタスク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。特に、大規模一般画像認識データセットである ImageNet [12] の学習済みモデルを 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 [13,14]。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。

## 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。

**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などを指す。病理画像解析におけるアプリケーションは、大きく分けて以下の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

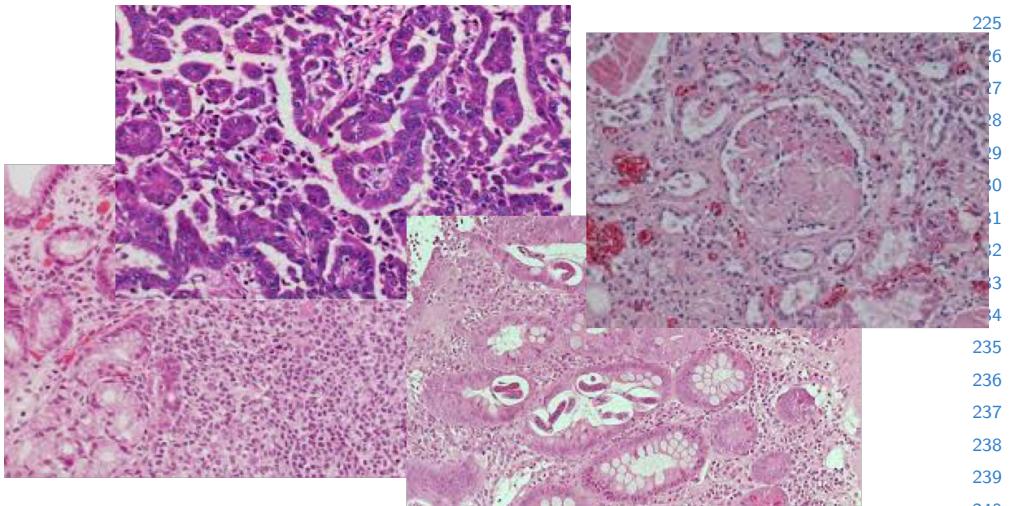


Fig. 3: 病理画像の例.

– 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

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 [5, 6]。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。

### 2.3 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高々  $1000 \times 1000$  以下である。数万ピクセル × 数万ピクセルに及ぶ WSI の全ピクセルを 1 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ものは WSI のわずか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ものを使用する。

### 2.4 使用される画像特徴量および識別器

画像認識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 などの hand-crafted の画像特徴量が同様に利用されていた [15]。また、病理画像は特定のオブジェクトが存在するか否かというよりも、画像全体の模様、もしくは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 [16–18]。

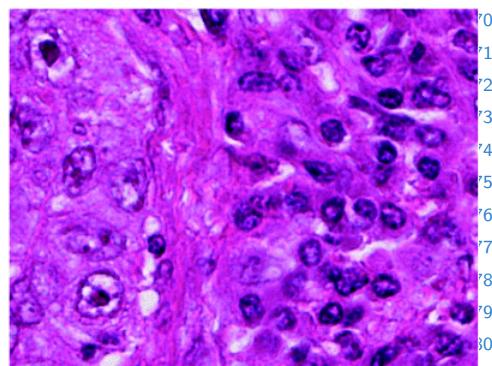
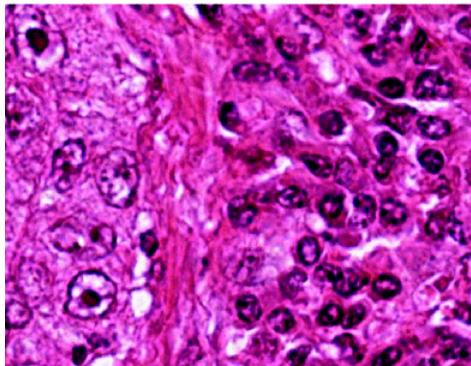


Fig. 4: 同じ細胞組織を異なる細胞スキャナーによって撮像したもの。色、解像度、輝度に大きな変化があることがわかる。

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 [19–23]。また CNN を利用する研究の中では、病変の検出率 [23] や癌の識別率 [20] が病理医に匹敵、あるいは超えるほどの精度を達成したという報告もある。

一般画像認識と比較して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 pretrained-model の中間特微量を抽出し識別器は SVM などを利用する場合や、2.1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、それらの中間特微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るこことが知られている [24, 25]

## 2.5 病理画像解析特有の性質

一般に医療画像解析に共通している特徴として、モデルが画像に対して獲得すべき不变性が多いことがある。特に病理画像解析では、各画像パッチは人手で切り取られたものであるので、モデルの予測は回転、位置に対しては不变であることが強く求められる。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（図 4）。このことから、色相、輝度に対する不变性が強く必要とされる。

## 2.6 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。

## 2.7 基本的な概念

能動学習 (Active Learning) [8] とは、機械学習における 1 つの枠組みである。一般に、教師つ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、低いアノテーションコストのもとでいかに高精度な予測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。

能動学習の基本的な流れを以下に示す

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. 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。以下それぞれのステップについて詳細を述べる。クエリ選択の詳細については次項で詳細を述べる。

クエリ問い合わせについて、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

巨大なラベル無しデータ集合 (サンプルプール) が予め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工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いる。病理画像解析ではラベルが付与されていない WSI およびそれに含まれる画像パッチは大量にある場合がほとんどであるため、本研究でも、ラベルなしサンプルプールが予め与えられている Pool-Based Sampling の問題設定を採用する。

再学習について、能動学習の研究では2通りの方法がある。1つは、クエリ問い合わせによってラベル付きデータが追加する度に、識別器のパラメータを初期値に戻してから再学習を行う方法である。ラベル付きデータが増える度にスクランチ学習を行うため、クエリ問い合わせ間隔が大きくなってしまう欠点がある。もう1つは、識別機のパラメータを初期値に戻すことなく、継続的に学習を行う方法である。ナイーブに学習を行うと、学習初期にラベルを付与されたデータの影響が大きくなってしまうため、新しいラベル付きデータに重みをつけて学習を行うなどの調整が必要となる。多くの研究では、再学習に時間のかからない線形識別器を用いているため、1つ目の再学習手法が採用される。識別器に CNN を用いた場合再学習に時間がかかるが、ハイパーパラメータ数の削減や学習の安定性という観点から、本研究でも再学習毎に識別器のパラメータを初期値に戻す方法を採用する。

## 360 2.8 クエリ選択の基準

361 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基準（クエリ選考基準）  
 362 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、  
 363  $\mathcal{U}$ 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  
 364 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。

$$365 \quad x^* = \operatorname{argmax}_{x \in \mathcal{U}} score(x) \quad (1)$$

366 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  
 367  $y$  の予測確率を  $P_\theta(y|x)$  とする。

371 **Uncertainty Sampling** [26] 最もシンプルな戦略で、モデルにとって予測が最  
 372 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、  
 373 ラベルの確率分布  $p(y|x)$  を出力する識別器であればいいという緩い条件のため幅広  
 374 く利用されている。曖昧性の定量評価として、事後確率最大ラベルの確率が小さい  
 375 サンプルを選択する Least Confident (式 2)，事後確率最大ラベルの確率と、その  
 376 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 (式  
 377 3)，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  
 378 (式 4) などが提案されている。

$$379 \quad score(x) = 1 - P_\theta(\hat{y}|x) \quad (2)$$

$$381 \quad score(x) = -(P_\theta(\hat{y}_1|x) - P_\theta(\hat{y}_2|x)) \quad (3)$$

$$384 \quad score(x) = - \sum_i P_\theta(y_i|x) \log P_\theta(y|x) \quad (4)$$

386 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、2値分類の場合は  
 387 全て等価な戦略となる。

389 **Query-By-Committee** [27] モデルの仮説空間内で、現在のラベル付きデータ  
 390 に対して Consistent な領域をバージョン空間と呼び、バージョン空間が大きいほど  
 391 パラメータの推定の分散が大きいことを示す。そこで、バージョン空間を縮小させる  
 392 サンプルがモデルの更新に寄与するはずだとする仮説に基づく戦略がある。  
 393 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  
 394 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  
 395 が Query-By-Committee (QBC) アルゴリズム [27] である。バージョン空間を近  
 396 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  
 397 モデル (Committee  $C = \{\theta^{(1)}, \dots, \theta^{(C)}\}$ ) を保持し、Committee 内のメンバの  
 398 予測の不一致度が最も高いサンプルを選択するという戦略である。不一致度の定  
 399 量評価として、Vote Entropy [28]、Average Kullback Leibler Divergence [29] が  
 400 ある。Vote Entropy は、Committee 内のメンバがどのラベルだと予測（投票）し  
 401 たかのばらつきを量化したもので、次式で表される

$$403 \quad score(x) = - \sum_{j=1}^m \frac{V(y_j)}{|C|} \log \frac{V(y_j)}{|C|} \quad (5)$$

405  $V(y_j)$  はクラス  $j$  の得票数,  $|C|$  は Committee のサイズを表す.

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

$$409 \\ 410 \quad score(x) = -\frac{1}{|C|} \sum_{c=1}^{|C|} KL(P_{\theta^{(c)}} \| P_C) \quad (6) \\ 411$$

412  $P_{\theta^{(c)}}$  は  $c$  番目の Committee の確率分布,  $P_C$  は Committee 全体の平均の確率予測  
413 分布を表す.

416 **Expected Model Change [30] / Expected Error Reduction [31]** データ  
417 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  
418 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  
419 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のモデルの更新量が最大  
420 となるサンプルを選択するのが Expected Model Change [30] と呼ばれる戦略で,  
421 また, モデルの更新量ではなくモデルの汎化誤差の減少量を評価することを目的  
422 とした戦略が Expected Error Change [31] である. 多くのケースでは, 全サンプ  
423 ルに全通りのラベルを仮定し, モデルの再学習を行う必要があるため膨大な計算  
424 コストを要する手法である.

426  
427 **Variance Reduction [32]** Geman et al. [33] の結果から, モデルの期待汎化誤  
428 差は以下のように分解することが出来る.

$$430 \quad E_T[(\hat{y} - y)^2 | x] = E_T[(y - E[y|x])^2] + (E_{\mathcal{L}}[\hat{y}] - E[y|x])^2 + E_{\mathcal{L}}[(\hat{y} - E_{\mathcal{L}}[\hat{y}^2])] \quad (7) \\ 431$$

432  $E_{\mathcal{L}}[\cdot]$  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  
433 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  
434 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  
435 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.  
436 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  
437 汎化誤差を小さくする戦略が Variance Reduction [32] である. また, モデルパラ  
438 メータの推定量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  
439 るという統計的性質がある (Cramel-Rao の不等式).

$$441 \quad Var(\hat{\theta}) \geq \frac{1}{I(\theta)} \quad (8) \\ 442$$

444 すなわち, Variance Reduction は, モデルのパラメータのフィッシャー情報量を最  
445 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  
446 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  
447 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  
448 モデルに対しては, 上記の Expected Model Change / Expected Error Reduction  
449 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が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1つに Information Density [34] 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。

$$score'(x) = score(x) \times \left( \frac{1}{|\mathcal{U}|} \sum_{i=1}^{|\mathcal{U}|} s(x, x_i) \right)^{\beta} \quad (9)$$

$s(\cdot)$  はサンプル間の類似度を算出する関数、 $\beta$  は類似度をどれほど重視するかを調整する係数。

## 2.9 他の関連する事項

**サンプリングバイアス** 能動学習では、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が実際のデータ集合とは異なるという問題がしばしば起こる。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1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、2.8項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 と Klaus 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ことでクエリ内サンプルの Diversity を担保するという戦略を提案した [35]。また、Chen と Krause らは、クエリ選考基準を劣モジュ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ュラ最適化問題に帰着することができるとして示した [36]。この時、貪欲法によって、最適な組み合わせで得られる性能の  $1 - \frac{1}{e}$  近似を与えることが保証される。しかし劣モジュ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ものが多い。

## 2.10 関連する先行研究と本研究の位置づけ

## 2.11 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、全ての画像パッチからなる画像群の分布は当然ながら独立同一分布でないため、能動学習を適用する場合サンプリ

495 ゲバイアスの影響を強く受けやすい。NalisNik et al. は Uncertainty Sampling に  
 496 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 [37]。Doyle  
 497 et al. は線形識別器を多数保持する Query-By-Committee を利用して癌検知の識  
 498 別タスクを解こうと試みた [38]。また, Zhu et al. は画像パッチからテクスチャ特  
 499 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  
 500 で劣モジュ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  
 501 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 [39]。しかし,  
 502 これらの研究では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  
 503 おり, 精度の点で十分であるとは言えない。線形識別器の性能の限界は第 5 章の  
 504 予備実験で検証する。

## 506 2.12 深層能動学習

507 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  
 508 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  
 509 組み合わせた研究も行われている [40, 41]。これらのアプローチは, Restricted  
 510 Boltzmann Machine や Stacked AutoEncoder などの事前学習により必要なラベ  
 511 ル数を削減した後, Uncertainty Sampling によるクエリ選択を行うというシンプ  
 512 ルなものが多い。CNN のように膨大なパラメータ数を持つ識別器を採用する場  
 513 合, ??節で述べたクエリ選考基準の多くは計算コストの観点で実用的ではないこ  
 514 とが最大の要因である。

## 517 2.13 本研究の位置づけ

518 前項で述べたように能動学習において深層学習を利用する研究の多くは, Uncer-  
 519 tainty 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,  
 520 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  
 521 の確率分布のエントロピーを利用することが多い。しかし, 一般に, 多層ニュ-  
 522 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  
 523 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  
 524 を示すわけではないという問題がある。また, 予測が 0.5 であるのはサンプルそ  
 525 のものがノイズとして持つ不確かさなのか, モデルパラメータの不確かさによる  
 526 ものなのかの判断が单一の予測からは判断できない。

527 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  
 528 動学習に効率的に組み込むために, Query-By-Committee の考えを基に考案した  
 529 Query-By-Dropout-Predictions を提案する。多層ニューラルネットワークの識別  
 530 精度向上に寄与すると考えられるサンプルを選択するために, Dropout によって  
 531 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  
 532 して利用しそれらの不一致度を計算することで近似的にバージョン空間を縮小する  
 533 サンプルを選択するアプローチである。この手法を利用し, 識別精度を犠牲にす  
 534 ることなく, 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構築する。  
 535 また, 病理画像に能動学習を適用した先行研究は数千枚程度の画像パッチから數  
 536 百枚を選択してラベルを付与して学習する規模のもので, 本研究で採用する数百  
 537 万オーダー病理画像群を利用した能動学習の研究は著者の知る限り初めてである。

### 540 3 提案システム

#### 541 3.1 概要

542 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  
543 を犠牲にすることなくアノテーションコストを抑えるために能動学習を病理画像  
544 解析に取り入れるシステムを構築する。

545 ここで、本システムの要求機能の詳細を述べる。

- 546 1. 実用に足る識別精度を達成する予測モデルを獲得できること
- 547 2. アノテーションコストが出来る限り低いこと
- 548 3. 現実的な計算コストでクエリを選択可能なアルゴリズムであること
- 549 4. WSI に含まれる大量に存在するラベルなしデータを扱えること

550 病理医にも匹敵する精度を達成するため、本研究では識別器に Convolutional  
551 Neural Network (CNN) を採用する。また、少ないラベル数でも学習を収束させ  
552 るために ImageNet で学習済みのモデルの重みをもとに fine-tuning によって学習  
553 を行う。2.2 節で述べたように、一般画像認識のために学習されたモデルでも、医  
554 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、  
555 妥当だと考えられる。

556 前章で述べたように、多くの能動学習アルゴリズムは CNN のように膨大なパ  
557 ラメータ数を持つアルゴリズムに対して適用する場合計算コストが爆発する問題  
558 がある。さらに、前章で述べたように深層学習と能動学習を組み合わせた手法の  
559 多くが採用している Uncertainty Sampling は、不確かさを陽にモデル化しない  
560 CNN に対しては相性が良くないと考えられる。そこで本研究では、深層能動学習  
561 の新たな手法として、Query-By-Dropout-Predictions を提案する。不確かさを定  
562 義しない多層ニューラルネットワークのパラメータを更新するサンプルを効率的  
563 に見つけるために、多層ニューラルネットワークの学習において正則化の目的で  
564 使用される Dropout を利用して擬似的に Query-By-Committee を再現する。さら  
565 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への不变性の担保  
566 を考慮し、各 Committee の prediction 時にもランダムな Data Augmentation を  
567 利用することで、識別に有効なサンプルのみではなく不变性を確保するために有  
568 効であるサンプルをクエリとして選択する手法も提案する。

569 また病理画像は、WSI 内に多くの類似画像が存在するため、パッチ内能動学  
570 習を採用した場合にクエリ内に類似画像が多数含まれることで、無駄なアノテ  
571 ションが生じてしまう可能性が高い。そこで、クラスタリングを事前にを行うこと  
572 でクエリ内の情報重複問題を緩和することを考える。その際に用いる特徴量につ  
573 いて、病理画像パッチはテクスチャとして扱われることが多いという特徴を考慮  
574 し、妥当だと考えられるものを列挙した上で実験から選定する。

575 さらに、病理画像解析に能動学習を適用するためには WSI に含まれる大量に  
576 存在するラベルなしデータを扱う必要がある。そこで WSI を多数含む病理画像  
577 データセットに含まれる大量の画像パッチを最大限に有効活用するための実用的  
578 な方法を提案する。

579 以下の節では、それぞれについて詳細を説明する。

585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629

Fig. 5: 本研究で提案するシステムの概念図

### 3.2 Query by Dropout Predictions

**手法の詳細** 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して Uncertainty Sampling によってクエリ選択を行う場合は、この確率分布のエントロピーの大きさなどをを利用して不確かさを定量評価する。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リングしているわけではない。また、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では、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 (QBC) [27] がしばしば用いられる。このバージョン空間を全て保持するのは多くの識別器において不可能であるため、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) によって近似的に表現する事が多い。Committee がバージョン空間を近似するには、以下の条件を満たす必要がある。

- 現在のラベル付きデータ集合に対しては Consistent であること
- 未知データに対する予測に分散を持つこと

しかし、QBC を CNN に適用することを考えた場合、CNN のように計算コストの重いモデルを複数同時に保持し学習するのは、メモリ、計算コストの観点から現実的ではない。また、膨大なパラメータ数を持つ多層ニューラルネットワークの最適解の探索空間は非常に広く、有限個の Committee メンバーによってバージョン空間を近似するのは難しい。

ここで、多くの深層学習のアーキテクチャにおいて、正則化のために利用される 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の性質に着目する。Dropout を利用し十分に学習されたネットワークの部分ネットワークは、現在のラベル付きデータに対してはほぼ Consistent であると考えられる。

そこで、本研究では Dropout によってサンプリングされた部分ネットワーク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 Query-By-Dropout-Predictions を提案する (図 6)。つまり、本体のネットワーク  $\mathcal{M}$  から生

630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648 Fig. 6: 提案する Query-By-Dropout-Predictions の概要図。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。  
649  
650  
651  
652 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  
653 Kullback Leibler Divergence を計算する。  
654  
655  
656  
657

$$score(x) = -\frac{1}{C} \sum_{c=i}^C KL(P_{\theta(c)} || P_C) \quad (10)$$

658 アルゴリズムの詳細を Algorithm.1 に示す。これらの部分ネットワークが Committe  
659 ee のメンバとして利用されるための条件としてあげた 2 つの性質について実  
660 験で検証し、Uncertainty Sampling のみを利用する方法と比較する。  
661  
662

**推論時での Data Augmentation の利用** CNN を学習させる際、訓練時に画像  
663 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  
664 性能を向上させる工夫が用いられる。ラベル付きデータが少ない場合は特に過学  
665 習を防ぐため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。  
666 第 2 章で述べたように、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  
667 すべき不変性が多く存在する。特に病理画像解析では、回転、位置に対する不変  
668 性、色相、輝度に対する不変性を獲得しているのが望ましい。  
669

前項で、Dropout によって生成された部分ネットワークによる Committee の不  
670 一致度を測る手法を提案したが、本研究ではそれだけではなく、Data Augmentation  
671 を推論時にも利用することで、さらに効率的にモデルを更新するサンプルを選択す  
672 る手法を提案する。提案手法の動作原理のイメージ図を図 7 に示す。通常 Dropout  
673 はネットワークの後段の全結合層のみで使用されることが多い。すなわち、Dropout  
674 によってサンプリングされた Committee の各メンバは、CNN の畳み込み層から

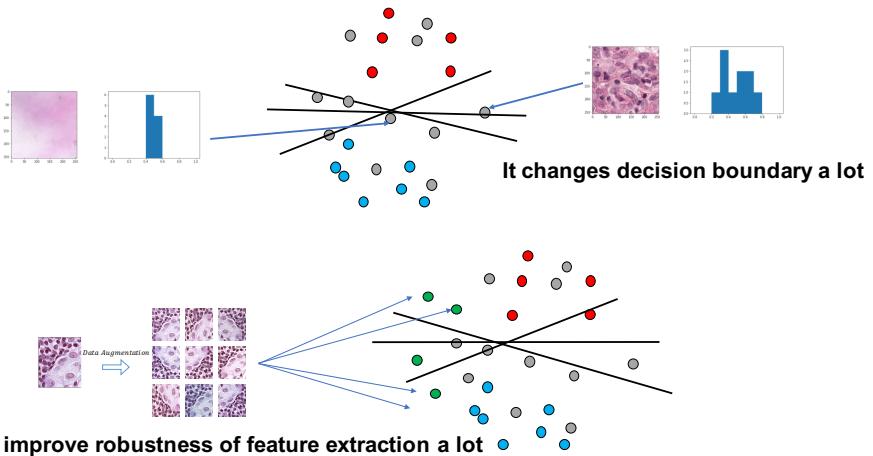


Fig. 7: 提案手法の動作原理のイメージ図。Dropout による不一致度の定量評価を Data Augmentation を併用することで効果を高める。

抽出された同一の特徴量に対する識別境界面の引き方が異なるのだと考えられる。この時、Committee の各メンバの予測の不一致度が高くなるようなサンプルは、Uncertain Sampling と比較して識別境界を大きく変更させるために有効であると考えられる。しかし CNNにおいては、識別境界の決定だけでなく特徴抽出の表現学習も重要な因子である。訓練時のみではなく不一致度の計算の際の推論時にも Data Augmentation を利用すると、同一のサンプルが特徴空間上の異なる点に配置されると考えられる。理想的にはそれらの点はすべて識別境界から見て同じ側に配置されるべきである。そこで、推論時の Data Augmentation を利用して予測が不一致したサンプルは、識別境界面を跨いでしまっているということであるので、そのサンプルを学習することで、特徴抽出が頑健になるのではないかと考えられる。

結局、推論時に Data Augmentation と Dropout を利用することで、識別境界の決定、特徴抽出の学習それぞれに有用であるサンプルが選択できるのではないかと考えられる。これを実験で検証する。

### 3.3 病理画像解析に適用するための工夫

**ラベルなしデータプールのサンプリング** 病理画像解析ではWSIを複数の画像パッチに分割して扱うが、一般画像認識で用いられるようなサイズ(256×256)だと仮定すると、1枚あたり数万～数十万枚の画像パッチを含むことになる。また、訓練時に使用できるWSIから全ての画像パッチを取り出しラベルなしデータセット $U$ として利用すると、その数は数百万オーダーとなる。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。しかし予め一定数をサンプリングして固定の $U$ を使用するのは、利用可能なデータ数を制限することになってしまう。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 $U_i$ を作成し、

---

**Algorithm 1** Query-By-Dropout-Predictions
 

---

```

720
721 Input: full network  $\mathcal{M}$ , unlabeled labeled dataset  $\mathcal{U}$ , active sampling size  $K$ ,
722 Output: query  $\mathcal{Q}$ 
723 1: sample Committee  $\mathcal{C} = \{\mathcal{M}_{p_1}, \mathcal{M}_{p_2}, \dots, \mathcal{M}_{p_c}\}$  from  $\mathcal{M}$  using dropout
724 2: for each  $x \in \mathcal{U}$  do
725   3:    $score(x) = disagreement\_score(x, \mathcal{C})$ 
726   4: end for
727   5:  $\mathcal{Q} \leftarrow \emptyset, \mathcal{D} \leftarrow \emptyset$ 
728   6: while  $len(\mathcal{Q}) < K$  do
729     7:    $x^* = \text{argmax}_{x \in \mathcal{U}_s} score(x)$ 
730     8:    $idx = cluster(x^*)$ 
731     9:   if  $idx \notin D$  then
732       10:     $\mathcal{Q} \leftarrow \mathcal{Q} \cup \{x^*\}, \mathcal{D} \leftarrow \mathcal{D} \cup \{idx\}$ 
733     11:   end if
734   12: end while
735   13: return  $\mathcal{Q}$ 
736

```

---

この中からモデル更新に寄与するサンプルを選択することで、現実時間内で最適なクエリを探索可能にしつつ、利用可能データを最大限に利用する方法を提案する。画像パッチをサンプリングする際は、各 WSI から均等に選択する。また、画像パッチの近傍は似た画像が多いことを考慮し、ある領域から 2 枚以上の画像パッチが選択されないように気をつけて行う。さらに、一度選択された領域は次のサンプリングでは避けるようとする。

**パッチ型能動学習への拡張** 2.6 節で述べたように、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、識別器を初期値に戻してから再学習させるのが一般的である。CNN を能動学習に利用する場合、一度の再学習に高い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パ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ラ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、本研究では予めラベルなしデータをクラスタリングし、一度のクエリでは同一のクラスターに属するサンプルを 2 つ以上選択しないことで、情報の重複を避ける。ここでは、クラスタリング手法はアルゴリズムのシンプルさから k-means に固定した。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

第 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、位置不变性と輝度変化への頑健性を有する Local Binary Pattern (LBP) [42] を採用する。実際には、回転不变性を担保するように 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

765 数の医療画像解析で利用されている。本研究では GoogleNet [43] の中間特徴  
 766 量を空間方向に平均を取ることで圧縮した 512 次元を利用した。  
 767

### Compact Bilinear Pooling による特徴量

768 近年では、深層学習を用いたテクスチャ解析に関する研究も盛んに行われて  
 769 いる。Bilinear Pooling [44] は、CNN の中間特徴量同士の相関を計算し空間  
 770 方向に平均を取ることで獲得される自己相関行列を特徴量として利用する手  
 771 法である。

$$G_{ij} = \sum_k F_{ik} F_{jk} \quad (11)$$

772  $G$  は Bilinear Pooling によって得られる自己相関行列、 $F$  は CNN の中間特徴  
 773 マップ、 $k$  は画像中の位置情報を表す。一般に CNN の特徴量次元（チャンネ  
 774 ル数）は 256 ~ 512 の大きな値であるため、直接計算した場合非常に高次元な  
 775 値となってしまう。そこで、ランダム行列によって Bilinear Pooling によって  
 776 得られる特徴量を近似する手法である Compact Bilinear Pooling (CBP) [45]  
 777 を本研究では採用する。ランダム行列の次元を増やすほど Bilinear Pooling の  
 778 近似精度が上がるという特徴になっている。CNN を用いたテクスチャ特徴量  
 779 として、GoogleNet [43] の中間特徴量を CBP によって 512 次元まで圧縮した  
 780 ものを採用する。

### 3.4 アルゴリズムの詳細

781 本研究で提案するアルゴリズムの詳細を以下に示す (Algorithm2)。

- 782 **step1** ラベルなしデータセット  $\mathcal{U}$ , 初期ラベル付きデータセット  $\mathcal{L}$  を準備する
- 783 **step2** CNN モデル  $\mathcal{M}$  に学習済みモデルの重みをセットする
- 784 **step3**  $\mathcal{L}$  を利用して,  $\mathcal{M}$  を fine-tuning により学習
- 785 **step4**  $\mathcal{U}$  から  $\mathcal{U}_i$  をサンプリングする
- 786 **step5** K-means によって  $\mathcal{U}_i$  を  $\mathcal{U}_{i_1}, \mathcal{U}_{i_2}, \dots, \mathcal{U}_{i_k}$  にクラスタリングする
- 787 **step6** Dropout により  $\mathcal{L}$  から Committee  $\mathcal{C} = \{\mathcal{M}_{p_1}, \mathcal{M}_{p_2}, \dots, \mathcal{M}_{p_c}\}$  を生成
- 788 **step7** Committee  $\mathcal{C}$  を利用し  $\mathcal{U}$  の disagreement score を計算
- 789 **step8**  $\mathcal{U}$  からスコアが高いものを, 同一クラスターからは 2 つ以上選択しないよ  
 うに選び, クエリ  $Q$  を作成
- 790 **step9**  $Q$  にラベルを付与し,  $\mathcal{L}$  に追加する
- 791 **step10** モデルの validation score が望みの値を達成するまで step3~step9 を繰  
 792 り返す

## 4 実験 : MNIST を用いた予備実験

### 4.1 概要

803 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  
 804 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  
 805 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  
 806 検証し、それらを Committee として扱いその予測の不一致度を利用する能够  
 807 学習のクエリ選考基準として有効であるかを確認する。

---

**Algorithm 2** Deep Active Learning for Pathological Image Analysis
 

---

810   **Input:** unlabeled dataset  $\mathcal{U}$ , initial labeled dataset  $\mathcal{L}$ , clustering size  $k$ , active sampling size  $K$ ,  
 811   **Output:** parameters of network  $\mathcal{M}$

812   1: Initialize pre-trained network  $\mathcal{M}_{pretrained}$   
 813   2: **repeat**  
 814   3:    $\mathcal{M} \leftarrow train(\mathcal{M}_{pretrained}, \mathcal{L})$   
 815   4:   sample partial unlabeled dataset  $\mathcal{U}_i$  from  $\mathcal{U}$   
 816   5:   Perform k-means clustering and devide  $\mathcal{U}_i$  to disjoint clusters  $\mathcal{U}_{i_1}, \mathcal{U}_{i_2}, \dots, \mathcal{U}_{i_k}$   
 817   6:    $\mathcal{Q} \leftarrow QBDP(\mathcal{M}, \mathcal{U}_i, K)$   
 818   7:   Query labels for  $\mathcal{Q}$   
 819   8:    $\mathcal{L} \leftarrow \mathcal{L} \cup \mathcal{Q}$   
 820   9: **until** performance is satisfactory

---

## 824   4.2 実験設定

825   **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  
 826   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  
 827   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  
 828   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  
 829   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。

830   **実験の詳細** 識別器には CNN を利用する。その構造を表 1 に示す。比較的小さな  
 831   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モデルの  
 832   パラメータを初期値に戻して再学習を行うことにする。ただし毎回ランダムに初  
 833   期化するのではなく、同じ初期値を利用することにする。これは、本来各 iteration  
 834   でクエリとして選択されるのはその時点でのモデルにとって有効なサンプルであ  
 835   り、パラメータを初期化するとそれらの重要度が変化してしまうという問題を少  
 836   しでも緩和するためである。

837   使用する CNN は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  
 838   ものの、ラベルを 1 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、  
 839   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  
 840   ためクラスタリングを行う。MNIST は画像 1 枚あたり 784 次元の比較的小さい  
 841   データであるため、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  
 842   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, 再  
 843   学習の epoch 数は 100 に設定した。一度に選択するクエリ  $Q$  のサイズは 5, 10,  
 844   50 の 3通りを実験した。また、クラスターの代表サンプルから無作為にサンプ  
 845   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  
 846   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 Random Crop  
 847   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  
 848   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 枚に到達するまで実  
 849   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  
 850   準偏差を計算した。

Table 1: MNIST の実験に使用した CNN の構造を示す。 (Keras の example プログラムを参考にした。)

layer	size	activation	dropout ratio
Convolution	$32 \times 4 \times 4$	relu	-
Convolution	$32 \times 4 \times 4$	relu	-
Max Pooling	$2 \times 2$	-	0.25
Fully Connected	128	relu	0.5
Fully Connected	128	relu	-

**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 以外はクラスタリングによるクエリの情報重複の回避を行った。

### Random Sampling

この実験のベースラインとなる選択基準。各クエリ問い合わせ毎にランダムに  $U$  から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 を使用せずに单一の予測分布のエントロピーを利用する。

$$score(x) = - \sum_i P_{\theta}(y_i|x) \log P_{\theta}(y|x) \quad (12)$$

### 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。不一致度には Committee の予測分布の Average KullbackLeibler Divergence を利用する。

$$score(x) = - \frac{1}{C} \sum_{c=i}^C KL(P_{\theta(c)} || P_C) - \sum_i P_C(y_i|x) \log P_C(y|x) \quad (13)$$

### 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なる。

## 4.3 予備実験

Query-By-Dropout-Predictions によって適切なクエリを選択できるためには、Dropout によって元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がそのバージョン空間を近似的に表現していなければならない。すなわち、各 Committee が訓練データに対しては Consistent であり、未知データに対してそれぞれの予測同士に分散を持つことが必要となる。この予備実験ではこれを示す。ランダムに

900  
901 Table 2: Dropout からサンプリングされた部分ネットワークの予測を計測した実  
902 験の結果

903 (a) 識別精度

	訓練データ	未知データ
Full Network	100.0 %	97.8 %
Committee 1	99.4 %	96.4 %
Committee 2	99.7 %	96.4 %
Committee 3	99.5 %	96.3 %
Committee 4	99.3 %	96.4 %
Committee 5	99.5 %	96.4 %
Committee 6	99.5 %	96.4 %
Committee 7	99.8 %	96.4 %
Committee 8	99.2 %	96.3 %
Committee 9	99.1 %	96.5 %
Committee 10	98.7 %	96.4 %

917 (b) 不一致度

	multiple CNN	Dropout network from single CNN
Average KL	$0.051 \pm 0.15$	$0.043 \pm 0.12$
Vote Entropy	$0.062 \pm 0.24$	$0.059 \pm 0.13$

923 (c) 不一致度

	multiple CNN	Dropout network from single CNN
Average KL	$0.061 \pm 0.19$	$0.051 \pm 0.12$
Vote Entropy	$0.071 \pm 0.20$	$0.063 \pm 0.19$

929  
930 1000 サンプルを選択しラベルを付与して学習を行った後、訓練データ、未知データ  
931 に対する Committee の各メンバの識別精度、それらの予測の不一致度をそれぞれ  
932 計算した。結果を表 2 に示す。

933 表から、Dropout からサンプリングされた部分ネットワークは訓練データに対  
934 しては Consistent であり未知データに対する予測はそれぞれ分散を持つことがわ  
935 かる。よって、これらの部分ネットワークを Query-By-Committee の Committee  
936 として利用するのは妥当であると考えられる。

#### 4.4 実験結果

本項では上述し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 8 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 9 に示す。提案した手法が 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。それぞれのクエリ選考基準を利用して構築された 1000 枚のラベル付きデータセットを学習した識別精度と、ラベルを全て使った場合の識別精度の比較した図を表 10 に示す。表より、提案手法を用いた場合、全てのラベルを利用して達成される識別精度に匹敵する性能を、それらの約 2% のラベル数で達成できていることがわかる。

また、ラベル付きデータセットが 100 枚の際に、各手法によってクエリとして選択されたサンプルを図 11 に示す。Random Sampling 以外の手法は、Random Sampling と比較すると認識が困難であると考えられるサンプルを選択する傾向にある。Uncertain Sampling ではクラスタリングを行っているものの、数字の 4 を含むクラスターは 10 以上あるため、偶然数字の 4 が極端に曖昧であったときに図のようなクエリになってしまうと考えられる。さらに、画像の中間特徴量を 2 次元に t-SNE によって射影し、Uncertain Sampling, QBDP によってそれぞれどのサンプルが選択されているかを表した（図 12）。提案手法は Uncertain Sampling と比較すると不確かなだけでなく、識別面を大きく変え得るサンプルを好む傾向にあることがわかる。

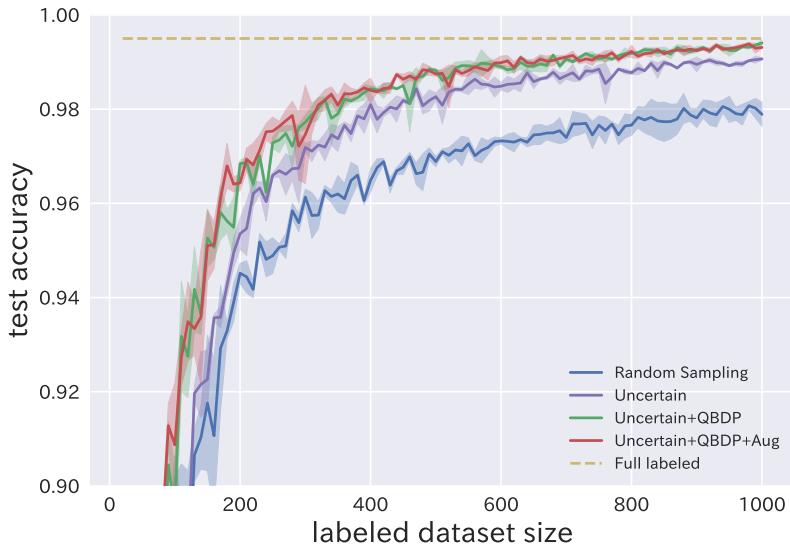


Fig. 8: 各手法を利用した場合のラベル付きサンプル数の増加に対するテスト精度の変化を示した図 ( $k = 10$ )

1035				1035
1036				1036
1037				1037
1038				1038
1039				1039
1040				1040
1041				1041
1042				1042
1043				1043
1044				1044
1045				1045
1046				1046
1047				1047
1048				1048
1049				1049
1050				1050
1051				1051
1052				1052
1053				1053
1054				1054
1055				1055
1056				1056
1057				1057
1058				1058
1059				1059
1060				1060
1061				1061
1062				1062
1063				1063
1064				1064
1065	Fig. 9: 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			1065
1066				1066
1067				1067
1068				1068
1069				1069
1070				1070
1071				1071
1072				1072
1073				1073
1074				1074
1075				1075
1076				1076
1077				1077
1078				1078
1079				1079

	90%	95%	98%
Random	110 ± 15	230 ± 22	900 ± 100
Uncertain Sampling	100 ± 5	170 ± 20	380 ± 25
Uncertain + QBDP	80 ± 5	140 ± 20	<b>290 ± 10</b>
Uncertain + QBDP + Aug	70 ± 15	140 ± 10	295 ± 30

(a)  $k = 5$ 

	90%	95%	98%
Random	110 ± 20	240 ± 10	900 ± 70
Uncertain Sampling	130 ± 5	180 ± 20	420 ± 30
Uncertain + QBDP	100 ± 10	150 ± 20	320 ± 10
Uncertain + QBDP + Aug	90 ± 0	150 ± 5	<b>300 ± 30</b>

(b)  $k = 10$ 

	90%	95%	98%
Random	150 ± 25	300 ± 25	900 ± 50
Uncertain Sampling	120 ± 0	170 ± 0	435 ± 25
Uncertain + QBDP	120 ± 0	170 ± 0	350 ± 25
Uncertain + QBDP + Aug	140 ± 25	190 ± 25	<b>320 ± 40</b>

(c)  $k = 50$

1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089

	識別精度	
Random	$98.1 \pm 0.3\%$	
Uncertain Sampling	$99.2 \pm 0.1\%$	
Uncertain + QBDP	$99.3 \pm 0.1\%$	
Uncertain + QBDP + Aug	<b><math>99.4 \pm 0.0\%</math></b>	
Full label (60000 label)	99.5 %	

(a)  $k = 5$ 

	識別精度	
Random	$97.9 \pm 0.3\%$	
Uncertain Sampling	$99.1 \pm 0.0\%$	
Uncertain + QBDP	<b><math>99.3 \pm 0.1\%</math></b>	
Uncertain + QBDP + Aug	<b><math>99.3 \pm 0.0\%</math></b>	
Full label (60000 label)	99.5 %	

(b)  $k = 10$ 

	識別精度	
Random	$98.0 \pm 0.1\%$	
Uncertain Sampling	$99.1 \pm 0.0\%$	
Uncertain + QBDP	<b><math>99.3 \pm 0.0\%</math></b>	
Uncertain + QBDP + Aug	<b><math>99.2 \pm 0.1\%</math></b>	
Full label (60000 label)	99.5 %	

(c)  $k = 50$ 

Fig. 10: それぞれのクエリ選考基準を利用して構築された 1000 枚のラベル付きデータセットを学習した識別精度と、ラベルを全て使った場合の識別精度の比較

1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169

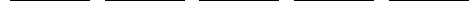
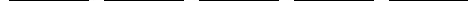
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169



(a) Random Sampling



(b) Uncertain Sampling



(c) Uncertain + QBDP

(d) Uncertain + QBDP + Aug

Fig. 11: 各手法によってクエリとして選択されたサンプルの例.

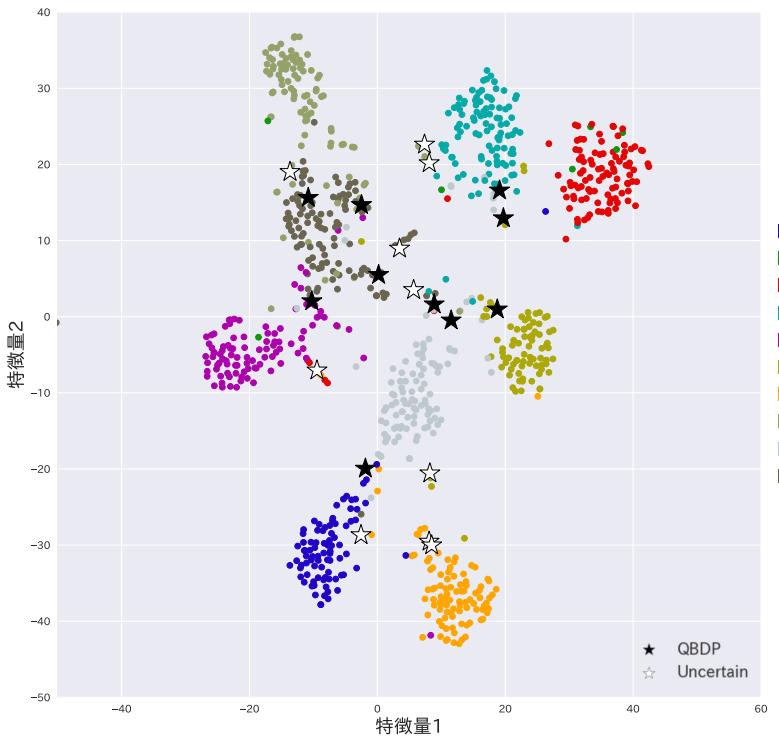


Fig. 12: サンプルの中間特徴量を2次元にt-SNEによって射影し、各手法によってどのサンプルが選択されているかを表した図。

## 1215 4.5 まとめ

1216 本研究で提案する Query-By-Dropout-Predictions を Mnist で検証し、その有効  
 1217 性を確認した。全てのラベルを利用した際の部分ネットワークが適切にバージョン  
 1218 空間を近似し、それを縮小させるサンプルを選択できていると考えられる。また、  
 1219 非常に簡単な Data Augmentation を追加することでわずかではあるが性能が  
 1220 向上したことから、病理画像での複雑な Data Augmentation を利用する場合は、  
 1221 さらに性能に変化が現れることが期待される。

## 1223 5 実験 2：病理画像データセットを用いた実験

### 1224 5.1 概要

1225 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  
 1226 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  
 1227 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  
 1228 験について説明する。

### 1229 5.2 実験設定

1230 **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  
 1231 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  
 1232 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  
 1233 したデータセットである（図.13）.. 各 WSI は一枚あたり  $100,000 \times 100,000$  ピ  
 1234 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。  
 1235 また、それら全ての画像パッチに対して癌か正常かの 2 値のラベルが付与されて  
 1236 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  
 1237 として問い合わせない限り与えられない状況とする。

### 1238 5.3 予備実験

1239 **クラスタリング手法の比較** 第 3 章で述べたクラスタリングに使用する特徴量を  
 1240 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  
 1241 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の平均を計  
 1242 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  
 1243 言える。データセットは 100,000 枚の病理画像からなり、癌と正常の割合は均等に  
 1244 調整した。表 3 に示すように、CNN を用いたテクスチャ特徴量である Compact  
 1245 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。

1246 **線形識別器との比較** 線形識別器による能動学習の研究は多数ある。ここで、特  
 1247 徴量を固定し線形識別器で学習した場合と特徴量も同時に学習した場合の精度を  
 1248 比較するために、病理画像データセットに対して実験を行った。識別機に用いる  
 1249 CNN は GoogLeNet [43] を採用した。一般画像認識で利用される数多くの CNN  
 1250 アーキテクチャの中でも比較的軽量で、医療画像解析でしばしば用いられるモデ  
 1251 ルであることから選択した。

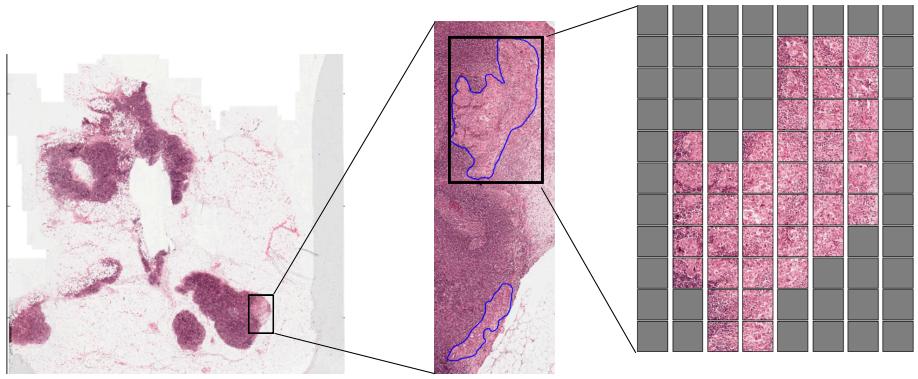


Fig. 13: Camelyon Dataset に含まれる WSI の例. 詳細なアノテーションが付与されている（青枠内部が癌）. 癌周辺の組織を  $256 \times 256$  のサイズの画像パッチに分割すると右の図のようになる.

Table 3: 比較実験の結果

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

図 4 に示すように、特徴量に深層学習によって得られたものを使用した場合でも、特徴量を同時に学習したものは線形識別器の精度を遙かに上回る結果となつた。また、通常の CNN よりも、Compact Bilinear Pooling によって空間の情報を落としてテクスチャ特徴量として扱ったほうが良い識別精度を達成することを確認した。

Table 4: 比較実験の結果

手法	Accuracy
LBP + 線形 SVM	80.2 %
CNN の中間特徴量 + 線形 SVM	84.4 %
Compact Bilinear Pooling + 線形 SVM	86.0 %
CNN + finetune	<b>94.1 %</b>
Compact Bilinear Pooling + finetune	<b>95.1 %</b>

## 1305 5.4 実験の詳細

1306 予備実験から、識別機に用いる CNN は GoogLeNet の中間特徴を Compact Bilinear  
 1307 Pooling によって圧縮し全結合層を接続する構造とした。以下に構造を示す（図  
 1308 14）。

1309 少ないラベルで高精度を達成するために、ImageNet の学習済みモデルを初期  
 1310 値として、fine-tuning によって学習を行う。

1311 訓練時に使用する WSI から全ての画像パッチを抽出すると約 5,000,000 枚に  
 1312 登る。これをラベルなしデータセット  $U$  とし、各クエリ問い合わせでは  $U$  からサ  
 1313 ンプリングされた  $U_i$  から選択する。 $U_i$  のサイズは 50,000 に設定した。

1314 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。  
 1315 予備実験より、クラスタリングに使用する特微量は Compact Bilinear Pooling を  
 1316 用いたテクスチャ特微量を採用する。Committee サイズは 10, k-means のクラス  
 1317 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。本実験  
 1318 では、実用的な状況に近づけるため、MNIST での実験のように各 iteration での  
 1319 学習回数を固定にせず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  
 1320 化しなくなるまで学習を続ける、という設定にした。バリデーションセットのサ  
 1321 イズは 100 とした。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  
 1322 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  
 1323 実験を開始した。各クエリ問い合わせ毎に 10,000 枚のテストデータに対する識別  
 1324 精度を計測し、ラベルを付与されたデータが 10,000 に到達するまで実験を続けた。  
 1325 比較手法は、以下の 3 つを使用した。

- 1326 – Random Sampling
- QBDP+Uncertain Sampling
- QBDP+Uncertain Sampling+Aug

1327 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  
 1328 した。

## 1333 5.5 実験結果

1334 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  
 1335 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  
 1336 化のグラフを図 15 に示す。学習済みモデルからの fine-tuning を行っているため、  
 1337 若干不安定ながら少ないラベルでも学習が進んでいることがわかる。Random な  
 1338 ベースラインと比較して、提案した手法は識別に有効なサンプルを選択できてい  
 1339 ることがわかる。さらに、MNIST の実験時よりも、推論時の Data Augmentation  
 1340 の利用の効果が顕著に現れていることがわかる。それぞれのクエリ選考基準を利  
 1341 用して構築された 1,000 枚のラベル付きデータセットを学習した識別精度と、ラベ  
 1342 ルを全て使った場合の識別精度の比較した図を表 5 に示す。表より、提案手法を用  
 1343 いた場合、全てのラベルを利用して達成される識別精度に遜色ない性能を、10,000  
 1344 枚のラベルで達成することができた。これは、実際ラベル付きデータがパッチレ  
 1345 ベルでは数百万枚存在することを考慮すると、100 分の 1 のラベル数で学習に成  
 1346 功したという事もできる。最後に、ラベル付きデータセットが 1,000 枚の際に、各  
 1347 手法によってクエリとして選択されたサンプルを図 16 に示す。

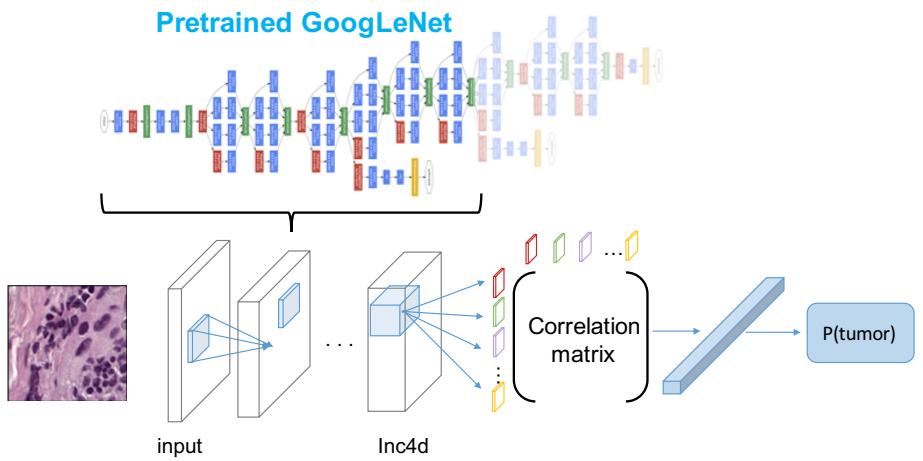


Fig. 14: GoogLeNet の特徴量を Compact Bilinear Pooling によって圧縮し、全結合層を接続した構造

Table 5: それぞれのクエリ選考基準を利用して構築された 1000 枚のラベル付きデータセットを学習した識別精度と、ラベルを全て使った場合の識別精度の比較

	AUC
Random	$0.959 \pm 0.002$
Uncertain + QBDP	$0.972 \pm 0.002$
Uncertain + QBDP + Aug	$0.975 \pm 0.002$
Full label ( $\simeq 6,000,000$ label)	0.98

## 5.6まとめ

本研究で提案するシステムを大規模病理画像データセットに適用し、システムの有効性を検証した。病理画像解析において CNN を用いてテクスチャ特徴量を利用しながら fine-tuning することで、他の手法よりも性能が良いことを実験で示した。巨大なパラメータを持つ CNN を採用し、識別精度を犠牲にすることなく、WSI 全体の画像パッチ群と比較した場合アノテーションコストを 1% 程度まで減らすことに成功した。また、それらを現実的な計算コストでの実現を可能とした。

## 6 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、能動学習を採用した。また、近年の画像認

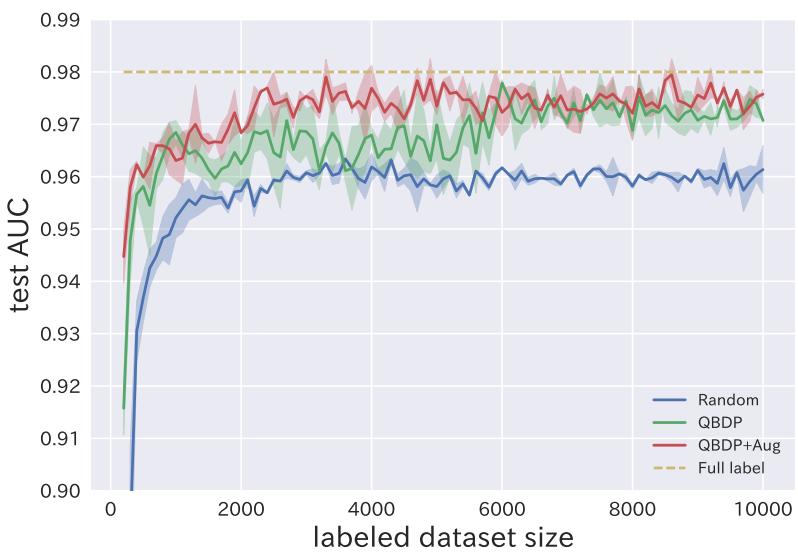


Fig. 15: 各手法を利用した場合のラベル付きサンプル数の増加に対するテスト精度の変化を示した図

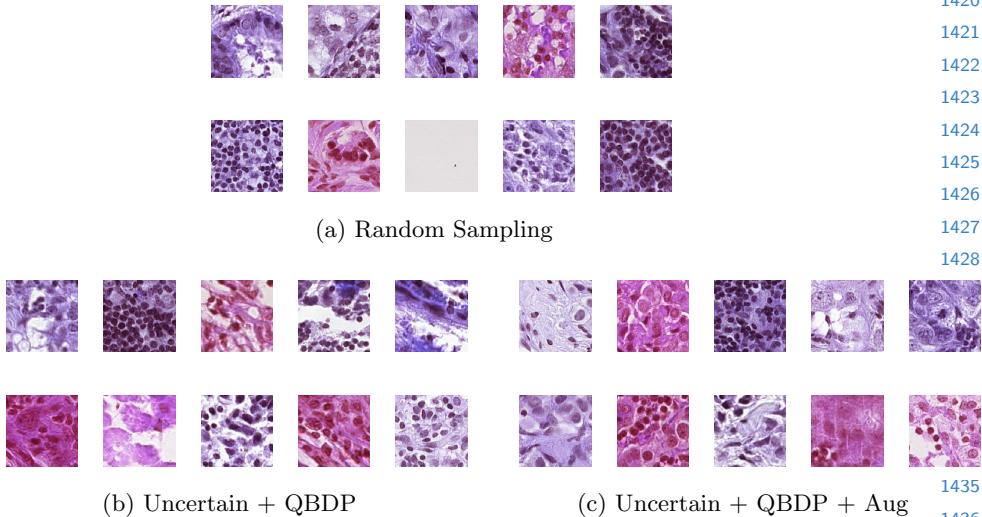


Fig. 16: 各手法によってクエリとして選択されたサンプルを示す。

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ための新たなクエリ選考基準 Query-By-Dropout-Predictions 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 によって元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によって近似的に再現することで、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である MNIST データセットで確認した。さらに、QBDP を含む提案システムを大規模病理画像データセットに対して適用し、アノテーションコストを抑えつつ高精度な予測モデルを獲得した。また、WSI に含まれる大量の画像パッチを能動学習の枠組みで効率的に扱うために、ラベルなしデータセットを各クエリ問い合わせ毎に一部サンプリングすることで利用可能なデータを制限しそぎることなく現実的な時間で有効なサンプルを選択する方法を提案した。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。

- Dropout によって元のネットワークからサンプリングされたネットワークは、Query-By-Committee の Committee として働くことがわかった。
- 深層能動学習において、提案した Query-By-Dropout-Predictions は従来の Uncertain Sampling と比較して少ないラベル付きデータで高精度な予測モデルを獲得できることがわかった。
- また、提案した Query-By-Dropout-Predictions は現実時間内で動作可能である。
- 不一致度を計算する推論時に Data Augmentation を使用することで、わずかに使用しない場合よりもモデルにとって良いサンプルを選択することが出来る。
- MNIST データセットでは、必要な 1,000 サンプルにラベルが付与されていれば、全てのデータにラベルがつけられている場合に匹敵する精度を達成できる。
- 病理画像解析に提案した深層能動学習を使用することで、アノテーションコストを大きく削減することが出来る。

## 6.2 今後の課題

**ラベルなしデータの活用** 本研究では、モデルのバージョン空間を縮小せるラベル付きデータセットを作成するための実験設定になっており、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れば良いのではないかと考えて実験を行った。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き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。

**継続的な fintune 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する方針を採用した。しかし、新たに追加されるラベル付きデータは、その時点でのモデルパラメータにとって最適なものであるため別のパラメータに初期化してしまった場合は真に最適となっているかは自明ではなくなる。また、再学習にはその分時間もかかるため、継続的な fintune をするための工夫を考える余地は残されている。

### 1485    6.3 Paper length

1486    Papers submitted for review should be complete. The length should match that  
1487    intended for final publication. Papers accepted for the conference will be allo-  
1488    cated 14 pages (plus references) in the proceedings. Note that the allocated 14  
1489    pages do not include the references. The reason for this policy is that we do not  
1490    want authors to omit references for sake of space limitations.

1491    Papers with more than 14 pages (excluding references) will be rejected with-  
1492    out review. This includes papers where the margins and formatting are deemed  
1493    to have been significantly altered from those laid down by this style guide. The  
1494    reason such papers will not be reviewed is that there is no provision for su-  
1495    pervised revisions of manuscripts. The reviewing process cannot determine the  
1496    suitability of the paper for presentation in 14 pages if it is reviewed in 16.

### 1497    6.4 Paper ID

1500    It is imperative that the paper ID is mentioned on each page of the manuscript.  
1501    The paper ID is a number automatically assigned to your submission when  
1502    registering your paper submission on CMT.

### 1504    6.5 Line numbering

1506    All lines should be numbered, as in this example document. This makes reviewing  
1507    more efficient, because reviewers can refer to a line on a page. If you are preparing  
1508    a document using a non-L<sup>A</sup>T<sub>E</sub>X document preparation system, please arrange for  
1509    an equivalent line numbering.

### 1512    6.6 Mathematics

1513    Please number all of your sections and displayed equations. Again, this makes  
1514    reviewing more efficient, because reviewers can refer to a line on a page. Also, it is  
1515    important for readers to be able to refer to any particular equation. Just because  
1516    you didn't refer to it in the text doesn't mean some future reader might not need  
1517    to refer to it. It is cumbersome to have to use circumlocutions like "the equation  
1518    second from the top of page 3 column 1". (Note that the line numbering will  
1519    not be present in the final copy, so is not an alternative to equation numbers).  
1520    Some authors might benefit from reading Mermin's description of how to write  
1521    mathematics: [www.pamitc.org/documents/mermin.pdf](http://www.pamitc.org/documents/mermin.pdf).

## 1524    7 Blind review

1525    Many authors misunderstand the concept of anonymizing for blind review. Blind  
1526    review does not mean that one must remove citations to one's own work. In fact  
1527    it is often impossible to review a paper unless the previous citations are known  
1528    and available.

Blind review means that you do not use the words “my” or “our” when citing previous work. That is all. (But see below for technical reports).

Saying “this builds on the work of Lucy Smith [1]” does not say that you are Lucy Smith, it says that you are building on her work. If you are Smith and Jones, do not say “as we show in [7]”, say “as Smith and Jones show in [7]” and at the end of the paper, include reference 7 as you would any other cited work.

An example of a bad paper:

An analysis of the frobncatable foo filter.

In this paper we present a performance analysis of our previous paper [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.

[1] Removed for blind review

An example of an excellent paper:

An analysis of the frobncatable foo filter.

In this paper we present a performance analysis of the paper of Smith [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.

[1] Smith, L. and Jones, C. “The frobncatable foo filter, a fundamental contribution to human knowledge”. Nature 381(12), 1-213.

If you are making a submission to another conference at the same time, which covers similar or overlapping material, you may need to refer to that submission in order to explain the differences, just as you would if you had previously published related work. In such cases, include the anonymized parallel submission [?] as additional material and cite it as

1. Authors. “The frobncatable foo filter”, BMVC 2014 Submission ID 324, Supplied as additional material `bmvc14.pdf`.

Finally, you may feel you need to tell the reader that more details can be found elsewhere, and refer them to a technical report. For conference submissions, the paper must stand on its own, and not *require* the reviewer to go to a techreport for further details. Thus, you may say in the body of the paper “further details may be found in [?]”. Then submit the techreport as additional material. Again, you may not assume the reviewers will read this material.

Sometimes your paper is about a problem which you tested using a tool which is widely known to be restricted to a single institution. For example, let’s say it’s 1969, you have solved a key problem on the Apollo lander, and you believe that the ECCV audience would like to hear about your solution. The work is a development of your celebrated 1968 paper entitled “Zero-g frobnication: How being the only people in the world with access to the Apollo lander source code makes us a wow at parties”, by Zeus.

You can handle this paper like any other. Don’t write “We show how to improve our previous work [Anonymous, 1968]. This time we tested the algorithm on a lunar lander [name of lander removed for blind review]”. That would be silly, and would immediately identify the authors. Instead write the following:

We describe a system for zero-g frobnication. This system is new because it handles the following cases: A, B. Previous systems [Zeus et al. 1968] didn't handle case B properly. Ours handles it by including a foo term in the bar integral.

...

The proposed system was integrated with the Apollo lunar lander, and went all the way to the moon, don't you know. It displayed the following behaviours which show how well we solved cases A and B: ...

As you can see, the above text follows standard scientific convention, reads better than the first version, and does not explicitly name you as the authors. A reviewer might think it likely that the new paper was written by Zeus, but cannot make any decision based on that guess. He or she would have to be sure that no other authors could have been contracted to solve problem B.

For sake of anonymity, it's recommended to omit acknowledgements in your review copy. They can be added later when you prepare the final copy.

## 8 Manuscript Preparation

This is an edited version of Springer LNCS instructions adapted for ECCV 2018 first paper submission. You are strongly encouraged to use L<sup>A</sup>T<sub>E</sub>X2 <sub>$\epsilon$</sub>  for the preparation of your camera-ready manuscript together with the corresponding Springer class file `l1necs.cls`.

We would like to stress that the class/style files and the template should not be manipulated and that the guidelines regarding font sizes and format should be adhered to. This is to ensure that the end product is as homogeneous as possible.

### 8.1 Printing Area

The printing area is 122 mm  $\times$  193 mm. The text should be justified to occupy the full line width, so that the right margin is not ragged, with words hyphenated as appropriate. Please fill pages so that the length of the text is no less than 180 mm.

### 8.2 Layout, Typeface, Font Sizes, and Numbering

Use 10-point type for the name(s) of the author(s) and 9-point type for the address(es) and the abstract. For the main text, please use 10-point type and single-line spacing. We recommend using Computer Modern Roman (CM) fonts, Times, or one of the similar typefaces widely used in photo-typesetting. (In these typefaces the letters have serifs, i.e., short endstrokes at the head and the foot of letters.) Italic type may be used to emphasize words in running text. Bold type and underlining should be avoided. With these sizes, the interline distance should be set so that some 45 lines occur on a full-text page.

**Headings.** Headings should be capitalized (i.e., nouns, verbs, and all other words except articles, prepositions, and conjunctions should be set with an initial capital) and should, with the exception of the title, be aligned to the left. Words joined by a hyphen are subject to a special rule. If the first word can stand alone, the second word should be capitalized. The font sizes are given in Table 6.

Table 6: Font sizes of headings. Table captions should always be positioned *above* the tables. The final sentence of a table caption should end without a full stop

Heading level	Example	Font size and style
Title (centered)	<b>Lecture Notes . . .</b>	14 point, bold
1st-level heading	<b>1 Introduction</b>	12 point, bold
2nd-level heading	<b>2.1 Printing Area</b>	10 point, bold
3rd-level heading	<b>Headings.</b> Text follows . . .	10 point, bold
4th-level heading	<i>Remark.</i> Text follows . . .	10 point, italic

Here are some examples of headings: “Criteria to Disprove Context-Freeness of Collage Languages”, “On Correcting the Intrusion of Tracing Non-deterministic Programs by Software”, “A User-Friendly and Extendable Data Distribution System”, “Multi-flip Networks: Parallelizing GenSAT”, “Self-determinations of Man”.

**Lemmas, Propositions, and Theorems.** The numbers accorded to lemmas, propositions, and theorems etc. should appear in consecutive order, starting with the number 1, and not, for example, with the number 11.

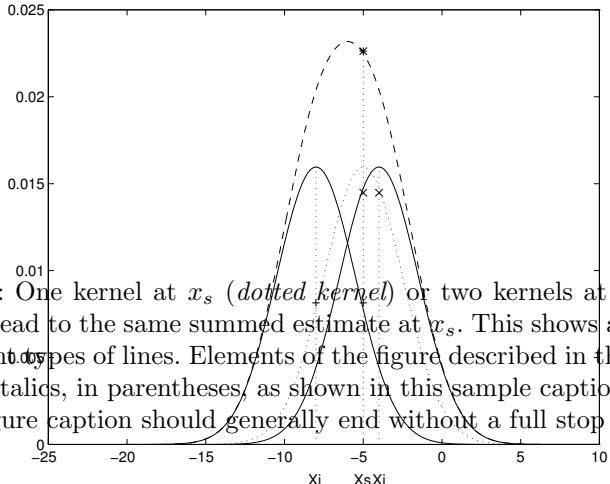
### 8.3 Figures and Photographs

Please produce your figures electronically and integrate them into your text file. For L<sup>A</sup>T<sub>E</sub>X users we recommend using package **graphicx** or the style files **psfig** or **epsf**.

Check that in line drawings, lines are not interrupted and have constant width. Grids and details within the figures must be clearly readable and may not be written one on top of the other. Line drawings should have a resolution of at least 800 dpi (preferably 1200 dpi). For digital halftones 300 dpi is usually sufficient. The lettering in figures should have a height of 2 mm (10-point type). Figures should be scaled up or down accordingly. Please do not use any absolute coordinates in figures.

Figures should be numbered and should have a caption which should always be positioned *under* the figures, in contrast to the caption belonging to a table, which should always appear *above* the table. Please center the captions between the margins and set them in 9-point type (Fig. 17 shows an example). The

1665 distance between text and figure should be about 8 mm, the distance between  
 1666 figure and caption about 5 mm.



1676  
 1677 Fig. 17: One kernel at  $x_s$  (*dotted kernel*) or two kernels at  $x_i$  and  $x_j$  (*left and*  
 1678 *right*) lead to the same summed estimate at  $x_s$ . This shows a figure consisting of  
 1679 different types of lines. Elements of the figure described in the caption should be  
 1680 set in italics, in parentheses, as shown in this sample caption. The last sentence  
 1681 of a figure caption should generally end without a full stop

1691 If possible (e.g. if you use L<sup>A</sup>T<sub>E</sub>X) please define figures as floating objects.  
 1692 L<sup>A</sup>T<sub>E</sub>X users, please avoid using the location parameter “h” for “here”. If you  
 1693 have to insert a pagebreak before a figure, please ensure that the previous page  
 1694 is completely filled.

## 1696 8.4 Formulas

1698 Displayed equations or formulas are centered and set on a separate line (with an  
 1699 extra line or halfline space above and below). Displayed expressions should be  
 1700 numbered for reference. The numbers should be consecutive within the contribu-  
 1701 tion, with numbers enclosed in parentheses and set on the right margin. For  
 1702 example,

$$1704 \psi(u) = \int_0^T \left[ \frac{1}{2} (\Lambda_0^{-1} u, u) + N^*(-u) \right] dt \quad (14)$$

$$1706 = 0? \quad (15)$$

1708 Please punctuate a displayed equation in the same way as ordinary text but  
 1709 with a small space before the end punctuation.

## 1710 8.5 Footnotes

1711 The superscript numeral used to refer to a footnote appears in the text either  
1712 directly after the word to be discussed or, in relation to a phrase or a sentence,  
1713 following the punctuation sign (comma, semicolon, or full stop). Footnotes should  
1714 appear at the bottom of the normal text area, with a line of about 2 cm in  $\text{\TeX}$   
1715 and about 5 cm in Word set immediately above them.<sup>1</sup>

## 1717 8.6 Program Code

1718 Program listings or program commands in the text are normally set in typewriter  
1719 font, e.g., CMTT10 or Courier.

1720 *Example of a Computer Program*

```
1723 program Inflation (Output)
1724   {Assuming annual inflation rates of 7%, 8%, and 10%,...
1725     years};
1726   const
1727     MaxYears = 10;
1728   var
1729     Year: 0..MaxYears;
1730     Factor1, Factor2, Factor3: Real;
1731 begin
1732   Year := 0;
1733   Factor1 := 1.0; Factor2 := 1.0; Factor3 := 1.0;
1734   WriteLn('Year 7% 8% 10%'); WriteLn;
1735   repeat
1736     Year := Year + 1;
1737     Factor1 := Factor1 * 1.07;
1738     Factor2 := Factor2 * 1.08;
1739     Factor3 := Factor3 * 1.10;
1740     WriteLn(Year:5,Factor1:7:3,Factor2:7:3,Factor3:7:3)
1741   until Year = MaxYears
1742 end.
```

1743 (Example from Jensen K., Wirth N. (1991) Pascal user manual and report. Springer,  
1744 New York)

## 1747 8.7 Citations

1748 The list of references is headed “References” and is not assigned a number in  
1749 the decimal system of headings. The list should be set in small print and placed  
1750 at the end of your contribution, in front of the appendix, if one exists. Please do

---

1751 <sup>1</sup> The footnote numeral is set flush left and the text follows with the usual word  
1752 spacing. Second and subsequent lines are indented. Footnotes should end with a full  
1753 stop.

not insert a pagebreak before the list of references if the page is not completely filled. An example is given at the end of this information sheet. For citations in the text please use square brackets and consecutive numbers: [?], [?], [?] ...

1755

1756

1757

1758

1759

1760

1761

1762

1763

1764

1765

1766

1767

1768

1769

1770

1771

1772

1773

1774

1775

1776

1777

1778

1779

1780

1781

1782

1783

1784

1785

1786

1787

1788

1789

1790

1791

1792

1793

1794

1795

1796

1797

1798

1799

Now we have reached the maximum size of the ECCV 2018 submission (excluding references). References should start immediately after the main text, but can continue on p.15 if needed.

## 1800 References

- 1802 1. Pantanowitz, L.: Digital images and the future of digital pathology. *Journal of pathology informatics* **1** (2010)
- 1803 2. Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B.: Histopathological image analysis: A review. *IEEE reviews in biomedical engineering* **2** (2009) 147–171
- 1804 3. Komura, D., Ishikawa, S.: Machine learning methods for histopathological image analysis. *CoRR* **abs/1709.00786** (2017)
- 1805 4. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. *arXiv preprint arXiv:1702.05747* (2017)
- 1806 5. Doyle, S., Agner, S., Madabhushi, A., Feldman, M., Tomaszewski, J.: 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*, IEEE (2008) 496–499
- 1807 6. Dundar, M.M., Badve, S., Bilgin, G., Raykar, V., Jain, R., Sertel, O., Gurcan, M.N.: Computerized classification of intraductal breast lesions using histopathological images. *IEEE Transactions on Biomedical Engineering* **58**(7) (2011) 1977–1984
- 1808 7. Geessink, O., Bárdi, P., Litjens, G., van der Laak, J.: Camelyon17: Grand challenge on cancer metastasis detection and classification in lymph nodes (2017)
- 1809 8. Settles, B.: Active learning literature survey. *University of Wisconsin, Madison* **52**(55–66) (2010) 11
- 1810 9. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* **115**(3) (2015) 211–252
- 1811 10. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580* (2012)
- 1812 11. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research* **15**(1) (2014) 1929–1958
- 1813 12. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database. In: *CVPR09.* (2009)
- 1814 13. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* (2014) 580–587
- 1815 14. Agrawal, P., Girshick, R., Malik, J.: Analyzing the performance of multilayer neural networks for object recognition. In: *European conference on computer vision*, Springer (2014) 329–344
- 1816 15. Caicedo, J.C., Cruz, A., Gonzalez, F.A.: Histopathology image classification using bag of features and kernel functions. In: *Conference on Artificial Intelligence in Medicine in Europe*, Springer (2009) 126–135
- 1817 16. Sertel, O., Kong, J., Lozanski, G., Shana'ah, A., Catalyürek, U., Saltz, J., Gurcan, M.: Texture classification using nonlinear color quantization: Application to histopathological image analysis. In: *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, IEEE (2008) 597–600
- 1818 17. Sertel, O., Kong, J., Catalyürek, U.V., Lozanski, G., Saltz, J.H., Gurcan, M.N.: Histopathological image analysis using model-based intermediate representations

- 1845 and color texture: Follicular lymphoma grading. *Journal of Signal Processing Systems* **55**(1-3) (2009) 169  
1846
- 1847 18. Nosato, H., Kurihara, T., Sakanashi, H., Murakawa, M., Kobayashi, T., Furuya,  
1848 T., Higuchi, T., Otsu, N., Terai, K., Hiruta, N.: An extended method of higher-  
1849 order local autocorrelation feature extraction for classification of histopathological  
1850 images. *IPSJ Transactions on Computer Vision and Applications* **3** (2011) 211–221  
1851 19. Hou, L., Samaras, D., Kurc, T.M., Gao, Y., Davis, J.E., Saltz, J.H.: Patch-based  
1852 convolutional neural network for whole slide tissue image classification. In: Pro-  
1853 ceedings of the IEEE Conference on Computer Vision and Pattern Recognition.  
1854 (2016) 2424–2433  
1855 20. Xu, J., Luo, X., Wang, G., Gilmore, H., Madabhushi, A.: A deep convolutional  
1856 neural network for segmenting and classifying epithelial and stromal regions in  
1857 histopathological images. *Neurocomputing* **191** (2016) 214–223  
1858 21. Litjens, G., Sánchez, C.I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs,  
1859 I., Hulsbergen-Van De Kaa, C., Bult, P., Van Ginneken, B., Van Der Laak, J.:  
1860 Deep learning as a tool for increased accuracy and efficiency of histopathological  
1861 diagnosis. *Scientific reports* **6** (2016) 26286  
1862 22. Chen, H., Qi, X., Yu, L., Heng, P.A.: Dcan: Deep contour-aware networks for  
1863 accurate gland segmentation. In: Proceedings of the IEEE conference on Computer  
1864 Vision and Pattern Recognition. (2016) 2487–2496  
1865 23. Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G.E., Kohlberger, T., Boyko, A., Venu-  
1866 gopalan, S., Timofeev, A., Nelson, P.Q., Corrado, G.S., et al.: Detecting cancer  
1867 metastases on gigapixel pathology images. *arXiv preprint arXiv:1703.02442* (2017)  
1868 24. Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D.D., Chen, M.: Medical image clas-  
1869 sification with convolutional neural network. In: Control Automation Robotics &  
1870 Vision (ICARCV), 2014 13th International Conference on, IEEE (2014) 844–848  
1871 25. Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway,  
1872 M.B., Liang, J.: Convolutional neural networks for medical image analysis: Full  
1873 training or fine tuning? *IEEE transactions on medical imaging* **35**(5) (2016) 1299–  
1874 1312  
1875 26. Lewis, D.D., Gale, W.A.: A sequential algorithm for training text classifiers. In:  
1876 Proceedings of the 17th annual international ACM SIGIR conference on Research  
1877 and development in information retrieval, Springer-Verlag New York, Inc. (1994)  
1878 3–12  
1879 27. Seung, H.S., Opper, M., Sompolinsky, H.: Query by committee. In: Proceedings of  
1880 the fifth annual workshop on Computational learning theory, ACM (1992) 287–294  
1881 28. Dagan, I., Engelson, S.P.: Committee-based sampling for training probabilistic  
1882 classifiers. In: Proceedings of the Twelfth International Conference on Machine  
1883 Learning, The Morgan Kaufmann series in machine learning,(San Francisco, CA,  
1884 USA) (1995) 150–157  
1885 29. McCallum, A., Nigam, K., et al.: Employing em and pool-based active learning  
1886 for text classification. In: ICML. Volume 98. (1998) 350–358  
1887 30. Settles, B., Craven, M., Ray, S.: Multiple-instance active learning. In: Advances  
1888 in neural information processing systems. (2008) 1289–1296  
1889 31. Roy, N., McCallum, A.: Toward optimal active learning through monte carlo esti-  
1889 mation of error reduction. ICML, Williamstown (2001) 441–448  
1890 32. Cohn, D., Atlas, L., Ladner, R.: Improving generalization with active learning.  
1891 *Machine learning* **15**(2) (1994) 201–221  
1892 33. Geman, S., Bienenstock, E., Doursat, R.: Neural networks and the bias/variance  
1893 dilemma. *Neural Networks* **4**(1) (2008)

- 1890 34. Settles, B., Craven, M.: An analysis of active learning strategies for sequence  
1891 labeling tasks. In: Proceedings of the conference on empirical methods in natural  
1892 language processing, Association for Computational Linguistics (2008) 1070–1079  
1893 35. Brinker, K.: Incorporating diversity in active learning with support vector  
1894 machines. In: Proceedings of the 20th International Conference on Machine Learning  
1895 (ICML-03). (2003) 59–66  
1896 36. Chen, Y., Krause, A.: Near-optimal batch mode active learning and adaptive  
1897 submodular optimization. In: ICML (1). (2013) 160–168  
1898 37. Nalisnik, M., Amgad, M., Lee, S., Halani, S.H., Vega, J.V., Brat, D.J., Gutman,  
1899 D.A., Cooper, L.A.: Interactive phenotyping of large-scale histology imaging data  
1900 with histomicsml. bioRxiv (2017) 140236  
1901 38. Doyle, S., Monaco, J., Feldman, M., Tomaszewski, J., Madabhushi, A.: An active  
1902 learning based classification strategy for the minority class problem: application to  
1903 histopathology annotation. BMC bioinformatics **12**(1) (2011) 424  
1904 39. Zhu, Y., Zhang, S., Liu, W., Metaxas, D.N.: Scalable histopathological image anal-  
1905 ysis via active learning. In: International Conference on Medical Image Computing  
1906 and Computer-Assisted Intervention, Springer (2014) 369–376  
1907 40. Wang, D., Shang, Y.: A new active labeling method for deep learning. In: 2014  
1908 International Joint Conference on Neural Networks (IJCNN). (July 2014) 112–119  
1909 41. Li, J.: Active learning for hyperspectral image classification with a stacked autoen-  
1910 coders based neural network. In: Image Processing (ICIP), 2016 IEEE International  
1911 Conference on, IEEE (2016) 1062–1065  
1912 42. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation  
1913 invariant texture classification with local binary patterns. IEEE Transactions on  
1914 pattern analysis and machine intelligence **24**(7) (2002) 971–987  
1915 43. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D.,  
1916 Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings  
1917 of the IEEE conference on computer vision and pattern recognition. (2015) 1–9  
1918 44. Lin, T.Y., RoyChowdhury, A., Maji, S.: Bilinear cnn models for fine-grained visual  
1919 recognition. In: Proceedings of the IEEE International Conference on Computer  
1920 Vision. (2015) 1449–1457  
1921 45. Gao, Y., Beijbom, O., Zhang, N., Darrell, T.: Compact bilinear pooling. In:  
1922 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.  
1923 (2016) 317–326  
1924  
1925  
1926  
1927  
1928  
1929  
1930  
1931  
1932  
1933  
1934