

能動学習のための教師なしアノテーション対象データ選択手法

渡邊正人[†] 菅原俊[†] 田口賢佑[†] 船津陽平[†]

概要：Convolutional Neural Network (CNN) を用いた教師あり学習に基づく Semantic Segmentation では、学習および評価用のデータへのアノテーションコストが課題になる。本研究では、能動学習に基づき、できるだけ学習効果の高いデータのみをアノテーションの対象とすることで、学習データへのアノテーションコストを低減するアプローチを検討する。学習済み CNN の認識精度が低い未学習データは、学習済み CNN にとって学習できていないと考えられ、学習効果は高いと思われる。しかし、認識精度を計算するためには、計算対象のデータの教師データが必要となる。本稿では、未学習データの学習効果の高さを、学習済み CNN の出力データを用いて算出し、アノテーション対象データを選択する手法を提案する。

キーワード：能動学習, Active Learning, データ選択手法, Semantic Segmentation, Uncertainty Sampling

1. はじめに

Convolutional Neural Network (CNN) を用いた教師あり学習による Semantic Segmentation では、大量の学習データを要する。大量の学習データの学習を実施するためには、学習データ入手コスト、アノテーションコスト、学習コストがかかる。学習データ入手コストとは、学習器に解かせたい問題において、サンプリングされ得る画像を入手するために要する時間を指す。アノテーションコストとは、教師データの無いデータに対して、教師データを作成するために要する時間を指す。学習コストとは、学習データを学習するために要する時間を指す。本研究は、学習データ入手コスト、アノテーションコスト、学習コストを含めた総合的なコストを低減することを目的としている。本稿では、学習データの量を減らすことで、アノテーションコストおよび学習コストを低減する方法を検討する。

教師あり学習を実施する学習器の認識精度は、学習器が解く問題によらず、学習データ量の Log スケールに比例する[1]。認識精度の向上には膨大な量の学習データが必要となり、学習データ入手コストおよびアノテーションコストを要する。単純に学習データの量を減らせば、学習器の認識精度が低下する可能性がある。

より少ない学習データで、より良い学習結果を得るための学習方法に、能動学習[2]がある。能動学習では、教師データの無い学習データから、認識精度の向上に最も寄与する学習データを、学習器に選択させる。選択させたデータに対して、人手でアノテーションが実施される。

能動学習を医療画像の Segmentation に適用した研究[3]や、Semantic Segmentation に向けた弱教師あり学習における弱教師の取得方法に適用した研究[4]、Segmentation Propagation に適用した研究[17]が報告されており、問題設定や適用先次第で効率的な学習が実現できる。本研究では、学習器に学習させる問題として、車載カメラ画像認識を扱

う。本稿での車載カメラ画像認識は、void, road, pedestrian, bicycle, vehicle, motorcycle の 6 クラスを認識対象のクラスとする Semantic Segmentation とする。車載カメラ画像の Semantic Segmentation は、Advanced Driving Assistant System(ADAS)や自動運転に向けた走行環境の認識の要素技術として期待されている。

車載カメラ画像認識では、認識対象のクラスのスケール、形状が多様である。よって、教師データには緻密さが必要だと予想される。本稿における車載カメラ画像認識は、弱教師は用いず、人手で作成された緻密な教師データを用いた学習による Semantic Segmentation を扱う。

車載カメラ画像認識では、認識を行う車載カメラから撮影される画像が非常に多様である。走行環境下に存在する認識対象物および認識対象外の背景などの組み合わせが膨大に存在する。さらに、逆光や暗所などの照度の変動、季節や気候の変動、カメラのレンズの変動など、様々な要因によって画像が変動する。

ADASや自動運転に Semantic Segmentation を応用するためには、実環境で起こり得る様々な要因にロバストな認識を実現する必要がある。ロバスト性を向上させるためには、一般的に大量の画像データが必要になる。ロバスト性の観点からも、能動学習によって学習データの量を減らすことは重要である。

能動学習におけるアノテーション対象データ選択手法には、様々な手法が存在する。問題設定にあったデータ選択手法を適切に設計するためには、学習器に学習させる問題に実際に適用し、結果を観察することが重要である。本稿では、車載カメラ画像の公開データセット Cityscapes[5]を用いた Semantic Segmentation のための学習に能動学習を適用し、学習データ量あたりの認識精度の推移を調査する。Cityscapes から少量の学習データセットを作成し、学習および認識精度の評価を実施する。学習データ量を増加させ、学習と認識精度の評価を実施した際の学習データ量の増加の方法と、認識精度の向上具合を観察する。

[†]京セラ株式会社 研究開発本部 システム研究開発統括部
ソフトウェア研究所 画像処理研究部

2. 関連研究

能動学習は、教師あり学習において、できるだけ少ない学習データで、学習器の認識精度を高精度化することを目的としている。能動学習のスキームを学習データセットに適用することで、その学習データセットによって得られる認識精度の向上効果を、できるだけ少ないデータ量で獲得することを目指す。

能動学習では、学習データの入手方法に応じた学習シナリオが提案されている。Pool-based Sampling では、未知なデータ集合（プール）が得られ、プールの中から学習に有用なデータに対してアノテーションを要求し、逐次的に学習することを繰り返す。Stream-based Selective Sampling では、単一の未知なデータに対して、アノテーションを要求するかどうかを決める。Membership Query Synthesis では、入力される可能性のある未知なデータに対して、アノテーションを要求する。Membership Query Synthesis における未知なデータには、学習器が学習によって仮定する入力データ空間において生成されたデータも含まれる。すなわち、人工的に生成されたデータも含まれる。

本稿では、Pool-based Sampling で未知データのプールが得られると仮定し、車載カメラ画像認識のための学習に能動学習を適用する。車載カメラで撮影された動画から車載カメラ画像を学習データとして抽出するといった作業は想像しやすく、プール単位で未知データが入手できるものとして、能動学習方法を検討する。

データ選択手法には、Uncertainty Sampling [6], Query-By-Committee[7], Density-Weighted Methods[8]がある。Uncertainty Sampling では、学習器による推論時に、推論結果が最も不確かな未知データに対して、アノテーションを要求する。Query-By-Committee では、複数の学習器を同時に学習させながら、学習器の推論結果がばらつく未知データに対してアノテーションを要求する。Density Weighted Methods では、学習器が仮定する入力データ空間における未知データの分布を考慮し、アノテーションを要求する。

本稿では、Uncertainty Sampling に基づくデータ選択手法を検討する。未知データを学習器に推論させた結果から、未知データに対する不確かさを推定し、不確かさの高いデータに対してアノテーションを要求する。

3. 提案手法

CNN を用いた教師あり学習に基づく画像認識では、認識精度の向上のために学習データが大量に必要になる。Semantic Segmentation のための学習データは、緻密な教師データが必要であり、アノテーションコストが膨大である。教師あり学習に基づく Semantic Segmentation のための学習に能動学習を適用し、学習データ量を低減することで、ア

ノテーションコストおよび学習コストの低減を狙う。

能動学習では、学習器の認識精度を向上させるデータをどのように推定するかが重要である。未知データを学習器に学習させる際、学習器の認識精度が良い未知データを学習させる場合に比べ、認識精度の悪い未知データを学習させる方が、学習器の認識精度の向上に寄与すると予想できる。認識精度が向上するかどうかは 4. 節で確認する。

しかし、認識精度の測定のためには教師データが必要になる。アノテーションを要求する対象のデータを選択するために、教師データを要すると、学習データ量を減らすことによるアノテーションコストの低減が実現できない。

3.1 不確かさ画像に基づくアノテーション対象データ選択手法

本研究では、認識精度の代替として不確かさ画像を用いて、学習器の認識精度を向上させるデータを推定する。不確かさ画像とは、Semantic Segmentation を行う対象のデータの不確かさを、画素単位で表現した画像である。8bit 階調の 1 チャンネル画像として表現されており、画素の推論結果が不確かであるほど画素値が 0 に近づく。

不確かさ画像は、Bayesian SegNet[9]と同様に、学習器に Dropout 層を実装することで取得可能にする。Dropout は過学習を抑える手法として使われるが、推論時に使用することで、隠れ層の構成が異なる出力を取得することができる。Dropout を用いてサンプリングした一定数の出力を用い、出力の画素ごとの平均を Semantic Segmentation 用の出力画像とし、出力の画素ごとの分散を不確かさ用の出力とする。

従来手法の Uncertainty Sampling では、データの推論結果の不確かさに基づいて、アノテーションを要求するか決める。Uncertainty Sampling における推論結果の不確かさを測る指標には Least Confident, Margin Sampling[10], Entropy[11]がある。Least Confident は、推論結果を出力した確率の低さを用いる指標である。Margin Sampling は、推論結果のクラスのうち、確率が 1 番目に高いクラスと、2 番目に高いクラスの確率の差を用いる指標である。Entropy は、推論結果のすべてのクラスの確率のエントロピーを用いる指標である。これらの指標は、画像分類のような、一つの画像データに対して一つのクラスを推論する問題設定において適用可能である。一方、Semantic Segmentation では、一つの画像データに対し、画素単位でクラスを推論するため、画素ごとに推論結果の不確かさは異なる。認識精度の悪いデータと認識精度の良いデータの不確かさ画像を観察し、観察結果をデータ選択手法に反映させ、データ選択手法の効果を確認する。

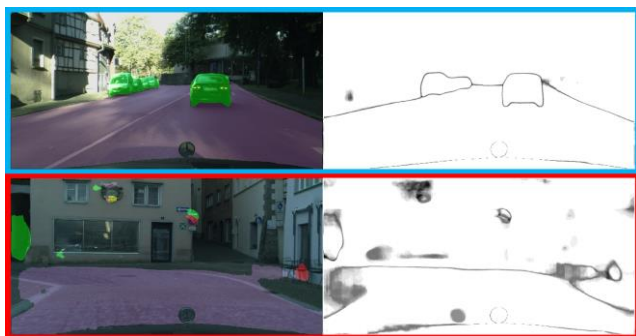


図 1. 推論結果画像と不確かさ画像. 上段は認識精度が最も高い画像の推論結果および不確かさ画像. 下段は認識精度が最も低い画像の推論結果および不確かさ画像.

図 1.に、観察対象の不確かさ画像の一例を示す. 図 1.は表 1.の実験条件で学習した学習器が、Cityscapes Validation データを推論した際の、推論結果画像および不確かさ画像である. 図 1.の不確かさ画像を観察すると、認識精度の悪いデータほど、不確かな画素の多い画像が出力される傾向があった. 不確かな画素とは、ここでは画素値が 255 でない画素を指す.

また、認識精度の悪いデータほど、不確かな画素の画素値の種類が多い傾向があった. 図 1.の下段を見ると、誤認識が起きている領域やその近傍に、様々な画素値の不確かな画素が存在することが確認できる.

認識精度の良いデータと悪いデータともに、クラスの領域間の境界において、不確かな画素が見られた. 同一クラスの領域間の境界では不確かな画素は見られず、異なるクラスの領域間の境界付近で不確かな画素が見られた. 推論時に Dropout 層を用いることと、CNN の量み込みの性質から、クラスの領域間の境界は、推論結果のクラスが変化する確率が高くなると考えられる. そのため、異なるクラスの領域間の境界では、不確かさが高くなる. 例えば、異なるクラス同士でオクルージョンが発生する場合は、不確かさの高い画素がよく見られる.

以上の観察結果に基づき、不確かな画素の画素数を、認識精度の向上度合いとするデータ選択手法(Number of Uncertainty Pixels: NUP)、不確かな画素の情報量を認識精度の向上度合いとするデータ選択手法(Entropy of Uncertainty Pixels: EUP)を提案する. 不確かな画素の情報量は、画素値に対する平均情報量によって求める.

NUP と EUP の差は、不確かさ画像の画素値の種類を考慮するかどうかである. NUP では、認識精度によらず、異なるクラス同士のオクルージョンが多発するような画像が、アノテーション対象のデータとして選択される懸念がある. EUP では、認識精度や異なるクラス同士のオクルージョンの発生数によらず、不確かな画素の種類が少なく、不確かな画素と確かな画素の出現頻度が同程度の時、アノテーシ

ョン対象のデータとして選択される懸念がある.

表 1. 予備実験の実験条件

項目名	内容
Train データ	Cityscapes Train データ 2975 枚
Validation データ	Cityscapes Validation データ 500 枚
データの解像度	幅 640px 高さ 320px
学習器に解かせる問題	6 クラス (void, road, pedestrian, bicycle, vehicle, motorcycle)の Semantic Segmentation
学習 Epoch 数	100
バッチサイズ	2
サンプリング数	30
Dropout connection 確率	0.25

表 2. 認識精度と不確かさの相関係数

	Number of Uncertainty Pixels	Entropy of Uncertainty Pixels
Cityscapes Train Mean Intersection over union	-0.472	-0.485
Cityscapes Validation Mean Intersection over union	-0.542	-0.549

3.2 認識精度と不確かさ画像の関係を確認するための予備実験

認識精度の代替として、不確かさ画像が活用できるか判断するため、学習済み学習器の認識精度と、不確かさ画像の関係を調査した. 実験条件を表 1. に示す. Cityscapes Train データ 2975 枚学習済みの学習器を用い、Train データおよび Validation データに対して、認識精度、不確かさ画像の不確かな画素数、不確かな画素の情報量を比較した. 本実験では、能動学習を適用せず、すべての学習データを表 1. の条件で一度に学習した.

実験に使用した学習器には、SqueezeNet [12]を Semantic Segmentation 向けに拡張し、Bayesian SegNet と同様に、Pooling 層の後段に Dropout 層を追加したものを用いた.

予備実験の結果を、図 2. 図 3. 図 4. および図 5. に示す. 図 2. 図 3. 図 4. および図 5. の横軸は認識精度を表している. 認識精度の評価指標には、mean Intersection over Union (mIoU)を用いた. 図 2. および図 4. の縦軸は不確かな画素の画素数を表している. 図 3. および図 5. の縦軸は、不確かな画素の情報量を表している. 図 2. および図 3.では、Cityscapes Train データに対する実験結果を示し、図 4. および図 5. では、Cityscapes Validation データに対する実験結果をそれぞれ示している. いずれの結果においても、縦軸と横軸に対して右肩下がりの散布図になっていることがわかる. 図 2. 図 3. 図 4. および図 5. のそれぞれの結果に対して、相関係数を求めた結果を表 2. に示す. いずれの場合においても、強い相関があるとは言いきれないものの、負の相関があることがわかる. 認識精度が高いほど不確かさは低く、認識精度が低いほど、不確かさは高い傾向が見て

取れる．この結果から，認識精度に基づくデータ選択手法

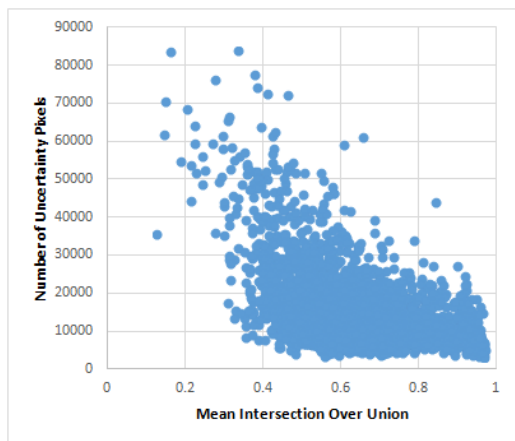


図 2. Cityscapes Train データにおける
認識精度と不確かな画素の画素数の散布図

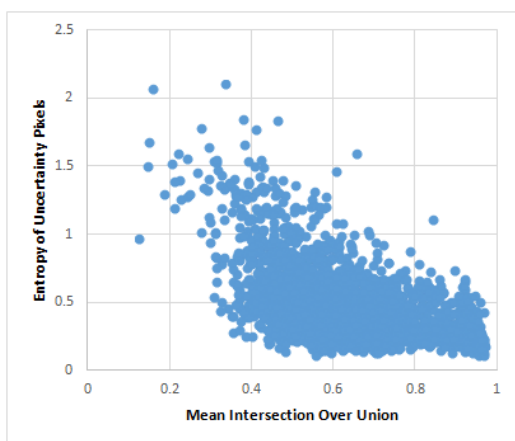


図 3. Cityscapes Train データにおける
認識精度と不確かさ画像の画素値の情報量の散布図

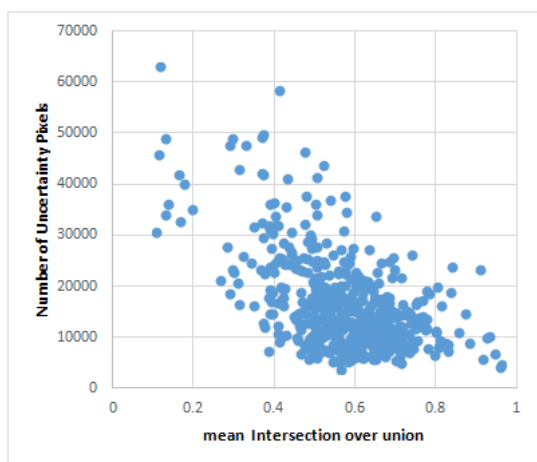


図 4. Cityscapes Validation データの
認識精度と不確かな画素の画素数の散布図

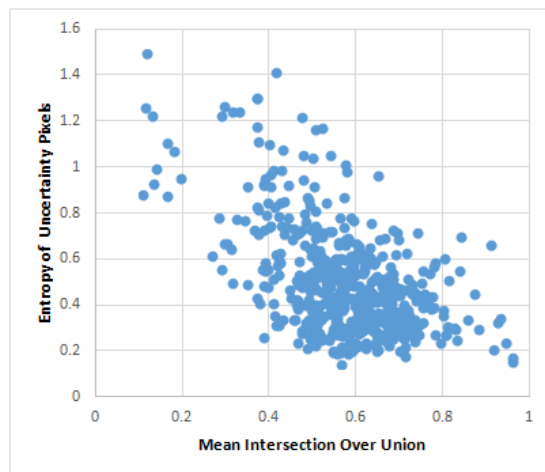


図 5. Cityscapes Validation データの
認識精度と不確かさ画像の画素値の情報量の散布図

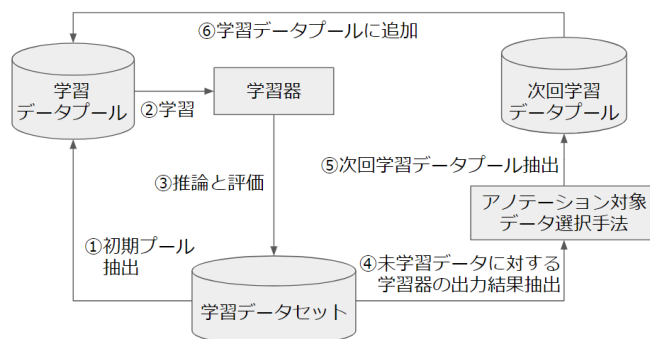


図 6. 能動学習の実験概要図

と似た効果が，NUP および EUP によって得られると予想できる．

4. 提案手法の効果確認と考察

手書き数字識別において，アノテーション対象のデータをランダムに決定した場合に比べ，能動的に決定した方が，少ないデータ量で識別精度が向上したことが報告されている[9]．以下では，車載カメラ画像を対象とした Semantic Segmentation においても，同様の傾向が見られるか確認する．

4.1 実験方法

図 6. に提案手法の効果を確認するための，能動学習の概要図を示す．本実験では学習データセットを少量のデータセットに分け，少量のデータセット単位で学習データ量を増やして学習した時の認識精度の推移を確認する．少量のデータセットを作る方法によって，少量の学習データ量で認識精度の向上が可能か確認する．教師ありアノテーション対象データ選択手法と比較して，認識精度の向上がどの程度効果的に実施できるか確認する．

表 3. 実験条件

項目名	内容
Train データ	Cityscapes Train データ 2975 枚
Validation データ	Cityscapes Validation データ 500 枚
データの解像度	幅 320px 高さ 160px
問題設定	6 クラス(void, road, pedestrian, bicycle, vehicle, motorcycle) の Semantic Segmentation
データ選択手法	Random, Inference Score, Number of Uncertainty Pixels, Entropy of Uncertainty Pixels
学習 Epoch 数	100
バッチサイズ	2
サンプリング数	30
Dropout connection 確率	0.25

実験は、5 種類の手順から成る 1 サイクルの作業を、6 回実施する。実験のサイクルを説明する。サイクルを開始する前に、初期プールを作成する必要がある。学習データセットから抽出した少量のデータセットを初期プールと呼ぶ。図 6. の①では、学習データセットからランダムに 500 枚、教師データ付きの学習データを選択し、初期プールを構築し、学習データプールとする。学習データプールとは、学習器に学習させる教師データ付きの学習データを指す。図 6. の②からサイクルが開始される。以下のサイクルを 6 サイクル実施する。図 6. の②では、学習データプールを学習器に学習させる。次に図 6. の③で、学習データプールを学習した学習器に、学習データセットを推論させ、認識精度を評価する。図 6. の④では、③での学習器の出力結果のうち、学習器が未学習なデータに対して出力した結果のみを抽出し、アノテーション対象データ選択手法への入力とする。アノテーション対象データ選択手法では、次回学習データプールを構成する学習データを選択する。本実験では、次回学習データプールのデータ量を 500 枚とした。Cityscapes Train データは合計 2975 枚なため、次回学習データプールのデータ量は、6 サイクル目のみ 475 枚とした。次回学習データプールは、学習データプールに追加され、1 サイクルが終了となる。以上の、図 6. における②から⑥までのサイクルを、6 サイクル実施する。

表 3. に実験条件を示す。提案手法と比較するデータ選択手法は、Random および Inference Score の 2 つである。Random は、学習器の出力によらず、未学習データからランダムに学習データを選択し、次回学習データプールとする。Inference Score は、学習器の推論結果画像を評価した結果、認識精度が悪いデータを未学習データから選択し、次回学習データプールとする。認識精度の評価指標として、mIoU を用いた。Inference Score では、認識精度の悪さを、認識精度の向上度合いとする。認識精度の計算のために教師データを用いる必要がある。そのため、Inference Score を教師ありアノテーション対象データ選択手法と呼ぶ。

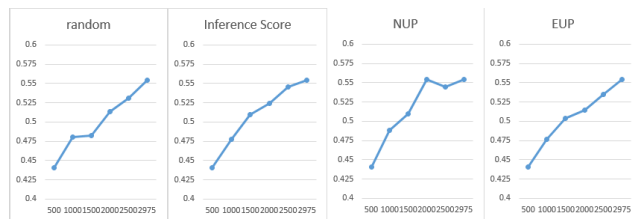


図 7. アノテーション対象データ選択手法ごとの Cityscapes Train データの認識精度の推移

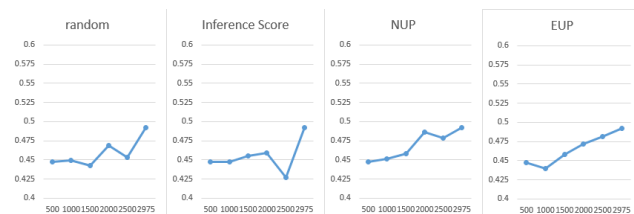


図 8. アノテーション対象データ選択手法ごとの Cityscapes Validation データの認識精度の推移

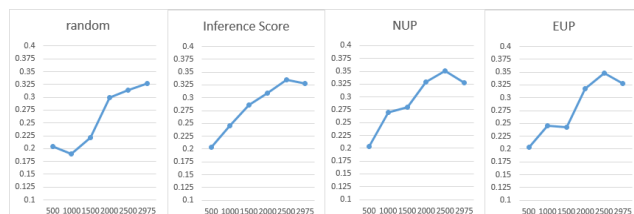


図 9. Cityscapes Train データの pedestrian に対する認識精度の推移

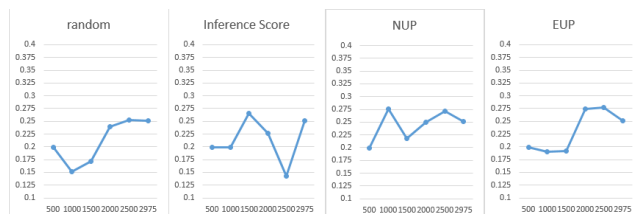


図 10. Cityscapes Validation データの pedestrian に対する認識精度の推移

NUP, EUP は認識精度の向上度の計算のために、教師データを必要としない。そのため、NUP, EUP を教師なしアノテーション対象データ選択手法と呼ぶ。

4.2 実験結果と考察

Cityscapes Train データに対する実験の途中経過を観察する。データ選択手法を Random および Inference Score とした場合の認識精度の向上具合を確認する。

図 7. は、データ選択手法ごとの Cityscapes Train データに対する認識精度の推移を、クラスごとに計測した結果である。図 8. は、データ選択手法ごとの Cityscapes Validation データに対する認識精度の推移を、クラスごとに計測した結果である。図 7. および図 8. いずれも、横軸を学習データプールのデータ量とし、縦軸をその学習データプールを

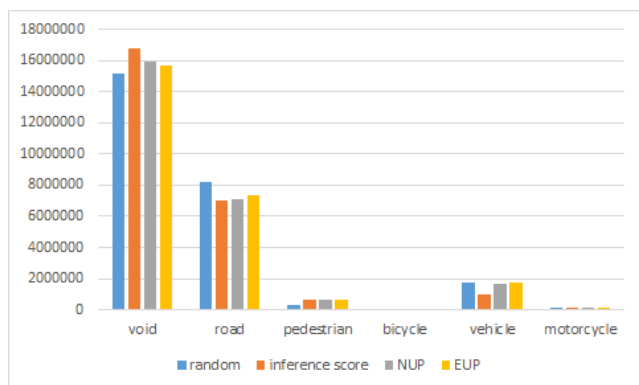


図 11. データ選択手法によって得られた学習データプールにおけるクラスの出現画素数

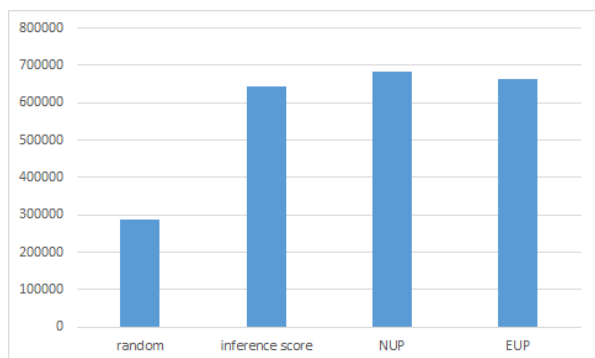


図 12. データ選択手法によって得られた学習データプールにおける pedestrian の出現画素数

学習した学習器の、Train データおよび Validation データそれぞれに対する mIoU 値とした。どの手法も、学習データ量を 500 枚から 1000 枚に増加させたときに、Train データに対する認識精度が 4% 上昇した。Validation データに対する認識精度は、ほぼ変わらない結果となった。

次にクラスを絞って認識精度を比較する。図 9, 図 10 では、pedestrian に対する認識精度の推移を、データ選択手法毎に比較している。図 9. および図 10. の横軸は学習データプールのデータ量とした。図 9. の縦軸は Train データを推論したときの mIoU 値とし、図 10. の縦軸は、Validation データを推論したときの mIoU 値とした。

1 サイクル目が終了した時点に着目すると、Random では、pedestrian に対する認識精度がわずかに減少した。Random によって選択された次回学習データプールの中に、pedestrian があまり含まれておらず、学習器が pedestrian の認識にとって不利な学習をした可能性がある。

一方で他の手法の場合では、pedestrian の認識精度が上昇している。Inference Score, NUP, EUP では、次回学習データプールに pedestrian が多く含まれ、pedestrian の認識精度の向上に影響したものと考えられる。

図 11. に、本実験における 1 サイクル目に得られた、次回学習データプールでの各クラスの出現画素数を示した。図 11. の横軸はクラス名、縦軸は画素数とした。void や road

に比べると、pedestrian の出現画素数は少なく、

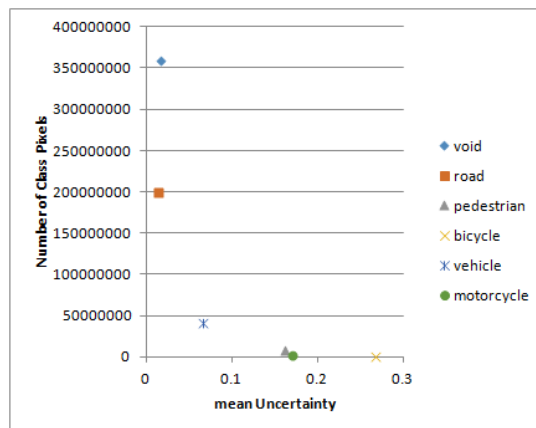


図 13. Cityscapes Train データの各クラスの画素数とクラスごとの不確かさの平均値の関係

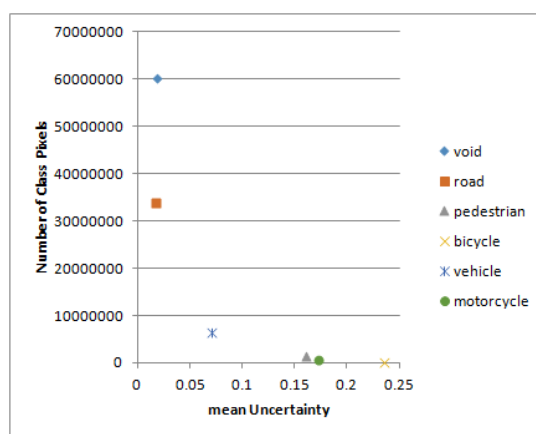


図 14. Cityscapes Validation データの各クラスの画素数とクラスごとの不確かさの平均値の関係

学習データ内に出現する頻度が低いことがわかる。

図 12. に、本実験における 1 サイクル目に得られた、次回学習データプールでの pedestrian の出現画素数を、データ選択手法ごとに示した。Random と比べ、Inference Score, NUP および EUP では、pedestrian が含まれるデータを、積極的にアノテーション対象のデータとして選択していることがわかる。以上から、提案手法には、pedestrian のような、学習データ内での出現頻度の低いクラスの認識精度を向上させる効果が期待できる。

学習データ量を減らし、アノテーションコストを下げるという目的では、データ選択手法は教師データが不要な手法である必要がある。Inference Score では、pedestrian が含まれるデータを、積極的に学習データとするために、教師データを要する。一方提案手法では、教師データなしに、pedestrian が含まれるデータを、積極的に学習データとすることができる。

4.3 今後の検討事項

どのデータ選択手法を適用する場合でも、初期プールの

選び方によって結果は変わる可能性がある。Random に学習データを選択する場合のほうが、提案手法よりも少ないデータ数で高い認識精度を達成する可能性は十分にある。

また、Cityscapes は、学習データとしての品質が非常によく、天候の変化や照度変化などの変動要因による画像データの変化が、非常に少ない特徴がある。そのため、データ選択手法を適用しない場合でも、高い認識精度が確認できたと考えられる。

提案手法では、不確かさ画像の画素値を、認識対象の全クラス平等に扱っている。しかし、不確かさの出現傾向は、認識対象のクラスをどの程度学習できているかによって変化すると考えられる。十分に学習できていないクラスは、クラスの推論確率が相対的に下がるはずであり、不確かさに影響を与えにくくなると予想できる。

表 1. に記載した予備実験用の Train データ、Validation データおよび予備実験の結果得られた推論結果画像を用い、認識対象の各クラスの画素数と、不確かさの関係を調査した。図 13. には、Cityscapes Train データの調査結果を、図 14. には、Cityscapes Validation データの調査結果を示した。どちらも縦軸にクラスの出現画素数を、横軸にクラス毎の不確かさの平均値を示している。不確かさの平均値は、不確かさ画像の画素値を bit 反転した後に正規化し、調査対象のデータすべてに対しての平均値を計算したものである。void や road のような、画像中に出現する頻度の高いクラスは、不確かさの平均値が低い。歩行者、自転車、バイクといった画像中に出現する頻度の低いクラスは、不確かさの平均値が高いことがわかる。不確かさ画像の画素値を用いて、Semantic Segmentation における未知データの不確かさを設計する際、認識対象のクラスの出現頻度を考慮する必要がある可能性がある。

能動学習を Semantic Segmentation に適用するにあたり、学習データプールのデータ量によっては、学習回数が増えちゃう懸念がある。学習回数の増加による学習コストの増加量と、アノテーションコストの削減量を比較する必要がある。また、本稿における提案手法の効果検証実験では、次回学習データプールを構成するデータ量を 500 枚で固定にした。認識精度の向上度合いや、学習済みの学習データプールのデータ量によっては、次回学習データプールのデータ量を可変にしても良い可能性がある。

学習データ入手コスト、アノテーションコストを低減する手法には、様々なアプローチが存在する。学習データ入手コストの低減には、人工的にデータを生成するアプローチ[12]がある。アノテーションコストの低減には、Amazon Mechanical Turk を活用するアプローチ[13]や、CG を活用するアプローチがある[14]。これらのアプローチを組み合わせることで、学習データ入手コスト、アノテーションコスト、学習コストを含めた総合的なコストを低減できると考えられる。

5. おわりに

車載カメラ画像の Semantic Segmentation を解く学習器に能動学習を適用し、学習データ量と認識精度の推移を観察した。教師ありアノテーション対象データ選択手法による能動学習を実施し、認識精度の悪いデータから学習することで、少ない学習データ量で、出現頻度の低いクラスの認識精度が向上することを確認した。学習器が出力する不確かさ画像を観察し、認識結果と不確かさ画像の関係を調査した。認識精度と不確かな画素数および不確かな画素の平均情報量の間には、負の相関がみられることがわかった。調査結果を教師なしアノテーション対象データ選択手法に反映させ、能動学習を実施し、その効果を検証した。出現頻度の低いクラスの認識精度の向上が見られた一方で、学習回数の増加に伴う学習コストの増加が懸念される。今後、認識精度の向上度合い、学習データ量、学習コストのバランスを加味し、CNN を用いた車載カメラ画像認識のための総合的な学習コスト低減を目指す。

参考文献

- [1] Sun, Chen, et al. "Revisiting unreasonable effectiveness of data in deep learning era." International Conference on Computer Vision (ICCV), pp. 843-852, 2017.
- [2] Settles, Burr. "Active learning." Synthesis Lectures on Artificial Intelligence and Machine Learning Vol. 6, No. 1, pp. 1-114, 2012.
- [3] Yang, Lin, et al. "Suggestive annotation: A deep active learning framework for biomedical image segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 399-407, 2017.
- [4] Vezhnevets, Alexander, Joachim M. Buhmann, and Vittorio Ferrari. "Active learning for semantic segmentation with expected change." Computer Vision and Pattern Recognition (CVPR), pp. 3162-3169, 2012.
- [5] Cordts, Marius, et al. "The cityscapes dataset for semantic urban scene understanding." Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3213-3223, 2016.
- [6] D. Lewis and W. Gale. A sequential algorithm for training text classifiers. In Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 3-12. ACM/Springer, 1994.
- [7] H.S. Seung, M. Opper, and H. Sompolinsky. Query by committee. In Proceedings of the ACM Workshop on Computational Learning Theory, pp. 287-294, 1992.
- [8] B. Settles and M. Craven. An analysis of active learning strategies for sequence labeling tasks. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1069-1078. ACL Press, 2008.
- [9] Kendall, Alex, Vijay Badrinarayanan, and Roberto Cipolla. "Bayesian SegNet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding." arXiv preprint arXiv:1511.02680 (2015).
- [10] Scheffer, Tobias, Christian Decomain, and Stefan Wrobel. "Active hidden markov models for information extraction." International Symposium on Intelligent Data Analysis, pp. 309-318, 2001.
- [11] Holub, Alex, Pietro Perona, and Michael C. Burl. "Entropy-based active learning for object recognition." Computer Vision and

- Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on. IEEE, pp. 1-8, 2008.
- [12] Iandola, Forrest N., et al. "SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size." arXiv preprint arXiv:1602.07360 (2016).
- [13] 櫻井啓介, 神谷祐樹, and 長谷川修. "競合型ニューラルネットワークを用いたオンライン準教師付能動学習手法." 電子情報通信学会論文誌 D 90.11 (2007): pp. 3091-3102.
- [14] Liu, Ming-Yu, Thomas Breuel, and Jan Kautz. "Unsupervised image-to-image translation networks." Advances in Neural Information Processing Systems. pp. 700-708, 2017.
- [15] Sorokin, Alexander, and David Forsyth. "Utility data annotation with amazon mechanical turk." Computer Vision and Pattern Recognition Workshops, pp.1-8, 2008.
- [16] Ros, German, et al. "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3234-3243, 2016.
- [17] Dutt Jain, Suyog, and Kristen Grauman. "Active image segmentation propagation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.2096-2106, 2016.