Vol.2018-CVIM-212 No.14
2018/5/10

オプティカルフローを用いた半教師あり学習にもとづくセマンティックセグメンテーション

水野 喬雄[1,a)]　鮫島 正樹[1]　菅野 裕介[1]　松下 康之[1]

概要：セマンティックセグメンテーションに代表されるような，学習データのラベル付けコストが非常に大きい様々なタスクにおいて，ラベルを自動生成できる合成画像を用いた研究が盛んに行われている．こうした中で，合成画像と実画像のギャップを改善するために，ラベル付き合成画像とラベル無し実画像を同時に用いた半教師あり学習によって精度を改善する方法が提案されている．本論文では，車載カメラ映像など合成画像と実画像の間でオプティカルフローが類似する場合が多々存在することに着目し，オプティカルフローにもとづく半教師あり学習を提案する．画像のみを用いた既存手法に比べて，局所的な動きの情報から予測したセマンティックラベルを拘束として用いることで，実画像に適応したセマンティックセグメンテーションを行う．合成画像に SYNTHIA データセットを，実画像に Cityscapes データセットを用いた評価実験によって，提案手法の有効性を示す．

Takao Mizuno[1,a)]　Masaki Samejima[1]　Yusuke Sugano[1]　Yasuyuki Matsushita[1]

1. 序論

画像中の物体やシーンを認識することはコンピュータビジョンにおいて重要な課題である．物体認識には粒度に応じて様々なタスクが存在し，画像中の物体がどのカテゴリに属するかを判別する画像分類 [1]，物体が画像中のどこにあるかを求める物体検出 [2]，画像をピクセル単位でクラス識別するセマンティックセグメンテーション [3] などが存在する．セマンティックセグメンテーションは自動運転の核となる技術として近年活発に研究されており，大規模なデータセット [4] が公開されて以来，ディープニューラルネットワーク (Deep Neural Network: DNN) を用いて特徴を抽出し，セマンティックセグメンテーションを行う方法が成功を収めている．一方，こうしたデータセットを作成する場合，実画像の収集は可能であっても，ピクセル単位でラベル付けを行う作業は大きな手間がかかる．

この問題に対して，コンピュータグラフィックス技術を用いて，様々な条件下での多量のラベル付き合成画像を自動的に生成し，学習データとして用いる研究が行われている [5]．しかし，合成画像と実画像のドメインは異なるため，合成画像で学習したセマンティックセグメンテーションネットワークを実画像の推定に用いても十分な性能を得ることはできない．そこで，ラベル付き合成画像とラベル無し実画像を同時に用いた半教師あり学習を行うことで精度を改善する試みがなされている．既存研究では，推定シーンが同じであると仮定して合成画像と実画像を同時に学習する手法 [6] や，正解ラベルのない実画像に対して推

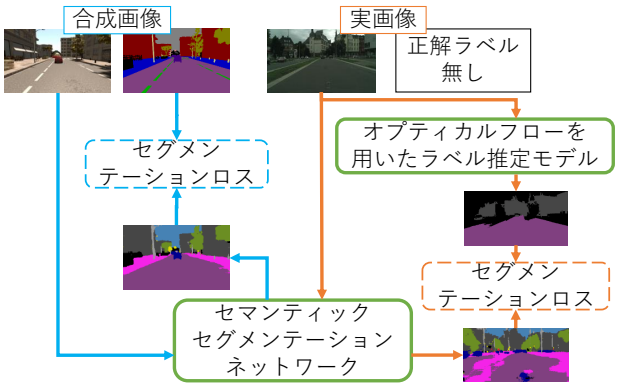


図 1: 提案手法の概要．オプティカルフローを用いたラベル推定モデルで実画像ラベルを推定し，推定した実画像ラベルと容易に自動生成できる合成画像のラベルを用いてセグメンテーションネットワークを学習することで，半教師有り学習を実現する．

[1] 大阪大学大学院情報科学研究科
[a)] mizuno.takao@ist.osaka-u.ac.jp

定したラベル分布を手がかりに学習する手法 [7] などが提案されている．しかし，車載カメラの場合では動きの情報もセマンティックなクラスに大きく関連するが，既存研究では考慮されていない．

本研究では，オプティカルフロー情報を利用して推定した実画像ラベルを用いてセマンティックセグメンテーションネットワークを学習する手法を提案する．車載カメラ映像において，例えば直進して撮影を行う場合，撮影物体は消失点から離れていくような動きをする．このような傾向は，車載カメラを想定した合成画像ならびに実画像において共通であり，ラベル無し実画像に対して合成画像のラベルを対応付ける手がかりとなりうる．図 1 に示す通り，提案手法では，正解ラベルの無い実画像に対してオプティカルフローを用いたラベル推定モデルで一度実画像のラベルを推定する．推定した実画像ラベルと容易に自動生成できる合成画像のラベルを用いてセグメンテーションネットワークを学習することで，半教師有り学習を実現する．評価実験により，提案手法がセマンティックセグメンテーションの性能向上に有効であることを示す．

2. 関連研究

本論文で対象とするセマンティックセグメンテーションについて 2.1 節，合成画像と実画像を利用するための半教師あり学習について 2.2 節で述べる．

2.1 セマンティックセグメンテーション

セマンティックセグメンテーションとは RGB 画像をピクセル単位でクラス識別を行うタスクであり，DNN を用いた研究が多くなされている．DNN の中でも畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) は，人間の脳の視覚野のニューロンの結合と似たニューラルネットワークであり，人間の認知と似た学習が行われると期待されており，実際に画像分類タスク等のコンピュータビジョンの分野において大きな成功を収めている．Long らは，画像分類タスクにおいて高いスコアを示しているネットワークにおける全結合層を，カーネルを用いて畳み込む Fully Convolutional Networks (FCN) を提案している [3]．FCN では，プーリングによりサイズが小さくなった最終出力である特徴マップのみをアップサンプリングしてセグメンテーション結果とすると，きめの粗い出力が得られるため，隠れ層の解像度が高い出力と組み合わせることで滑らかな出力を得ている．他にも多くのセマンティックセグメンテーションネットワークが提案されているが，新たな環境で充分な量の学習データを用意することは時間と労力を必要とする作業である．Cordts らは Cityscapes [8] データセットの作成において，良質なラベル付けをするのに画像一枚当たり 90 分の時間を費やしたと報告している．この問題に対して，近年コンピュータグラフィックスの進歩により，実画像に対して人手によるピクセル単位でのラベル付けを行わず，ラベル付きの合成画像を手に入れられるようになった．しかし，合成画像を使って学習したネットワークで実画像に対して推定を行うと，実画像で学習したネットワークと比べて精度は大幅に低下してしまう．

2.2 半教師あり学習

半教師あり学習とは，ラベル付きデータとラベル無しデータが混在するデータでモデルを学習させることであり，ラベル付きデータのみを用いて学習させたモデルよりも予測精度を向上させることが目標である．ここで，半教師あり学習で扱うデータがデータセットの違いなどにより学習データとテストデータが異なる母集団分布に従う場合，すなわち学習データとテストデータが異なるドメインに属する場合を対象とした半教師あり学習の研究が様々なタスクに対して行われている．その中でもセマンティックセグメンテーションの分野において Hoffman らは，異なるドメインであってもネットワークの中間層において共通の特徴空間を持つと仮定して，両ドメインの画像を学習する方法を提案している [6]．元ドメインのラベル付き画像に対しては，従来の手法と同様に，セマンティックセグメンテーションのロスをネットワークの学習に使う一方で，FCN においてアップサンプリングする前の特徴が，元ドメイン画像と目標ドメイン画像で近しくなるようにネットワークの重みを学習する．これに対して Zhang らは，異なるドメインにおいて共通の特徴空間を持つと仮定するのではなく，目標ドメインの情報を別のタスクで推定し，その情報をもとにセマンティックセグメンテーションネットワークを学習する手法を提案している [7]．推定する目標ドメインの情報の一つにラベル情報が含まれているが，推定には RGB 画像のみを用いており，ドメインに特徴的な情報を付加することによって精度を向上する余地があると考えられる．そこで本研究では，元ドメインも目標ドメインも車載カメラから撮られた映像である点に着目し，オプティカルフローを特徴として加える．

3. 提案手法

提案手法では，図 1 に示すように，セマンティックセグメンテーションネットワークに実画像に対応するラベルを学習させるために，ラベル付き合成画像だけでなくラベルが無い実画像も学習に加える．そのためには学習に用いる実画像のラベルを得る必要があり，オプティカルフローを用いたラベル推定モデルで一度大まかなラベル推定を行う．以下，3.1 節でオプティカルフローを用いた実画像ラベルの推定について述べ，3.2 節で実装の詳細について述べる．

3.1 実画像ラベルの推定方法

提案手法では，実画像ラベルの推定の際に RGB 画像

図 2: 合成画像と実画像のオプティカルフローの比較．白色は値が小さく，濃色ほど値が大きいことを表す．各色相の違いによって移動方向を示している．

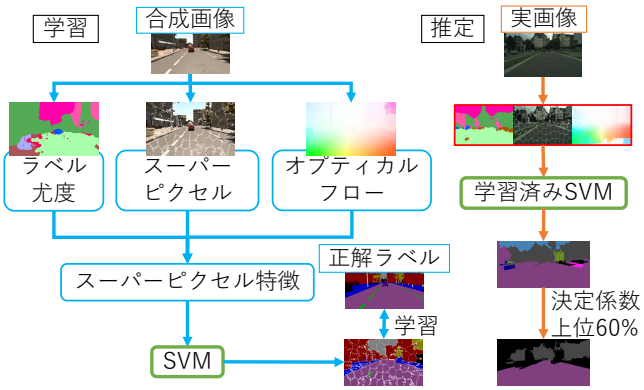


図 3: 実画像ラベルの推定方法．図 1 におけるオプティカルフローを用いたラベル推定モデルの学習方法と推定についての概要．

に加えてオプティカルフローを用いる．ここで，CNN を利用したオプティカルフロー推定ネットワークである Flownet2.0 [9] で推定した合成画像と実画像のオプティカルフローの値に応じて着色したものを図 2 に示し，オプティカルフローから得られる情報について説明する．図 2 が示すように，車載カメラで撮られた連続画像においては，背景クラスのオプティカルフローは消失点を中心にして画像外側に広がっていく．例えば，空や建物のクラスを持つピクセルは次のフレームでは画像中の上方に流れて，道路のクラスを持つピクセルは下方に流れていくことが多くなる．よって，局所的にラベルを推定する際，その場所のオプティカルフローが分かれば，画像中における位置が特徴として得られる可能性がある．また，自動車が直進する場面では，オプティカルフローは消失点を中心にして画像外側に広がると述べたが，右折するときはピクセルは全体的に左に流れ，左折するときは右に流れる．停止しているときは背景部分のピクセルは動かない．このように背景クラスの動きには一定の傾向がある．そこで，周囲は一定の動きをしているにもかかわらず，ある部分だけ移動の向きや大きさが異なれば，その部分が動物体であるという情報が得られる．

次に，図 1 におけるオプティカルフローを用いたラベル推定モデルの概要を図 3 に示す．ラベル推定は，画像における物体配置が合成画像と実画像で類似する点を考慮して，図 3 に示すように分割領域であるスーパーピクセル単位で行う．合成画像を用いてスーパーピクセル毎に，様々なクラスが含まれる異なるデータセットに対するセマンティックセグメンテーションスコアであるラベル尤度とオプティカルフローを特徴として，一つのラベルを出力する SVM を学習し，学習済み SVM を実画像に適用することで実画像ラベルの推定を行う．

3.2 実装の詳細

セマンティックセグメンテーションのネットワークは図 4 に示すように FCN-8s [3] の構造をもとにしたものを利用し，以下のセグメンテーションに関するロス L を逆伝搬することで学習する．

$$L = -\frac{1}{N}\sum_{k}^{K} \mathbf{t}_k{}^{\top} \log \mathbf{y}_k \qquad (1)$$

ここで，N は画像の総ピクセル数であり，K は分類クラス数である．$\mathbf{t}_k \in \mathbb{R}^N$ は k 番目のチャネルにおける真値のラベルを表現したものである．真値のラベルはそれぞれのピクセルが k 番目のクラスに属しているなら 1, そうでないなら 0 の値を持つ one-hot 表現を用いる．$\mathbf{y}_k$ は k 番目のチャネルにおける最終層の出力であり，各ピクセルでの値を全てのチャネルで和を取ると 1 となるように softmax 関数が施されている．よってロス L は，正解ラベルに対応するチャネルの出力が 1 に近ければ小さくなり，0 に近ければ大きくなる．また，ネットワークの学習にはミニバッチ学習を用いて，ミニバッチ中に含まれるデータ数は合成画像と実画像で同じとする．

次に，オプティカルフローを用いた実画像ラベルの推定における実装の詳細を説明する．図 3 において，スーパーピクセルに分割する際は Zhang ら [7] の手法と同様に，線形スペクトルクラスタリング [10] を用いて一枚の画像を約 100 個のスーパーピクセルに分割する．スーパーピクセルの特徴は，事前に PASCAL CONTEXT [11] データセットで学習した FCN-8s [3] に画像を入力して，ピクセル毎に 59 次元のスコアとなるラベル尤度が得られる．また，Flownet2.0 [9] を用いて各ピクセルにおける 2 次元のオプティカルフローを求め，元の 59 次元のスコアと結合することで，ピクセルごとに 61 次元の特徴ベクトルを得る．これをスーパーピクセル内で平均を取ることで，スーパーピ

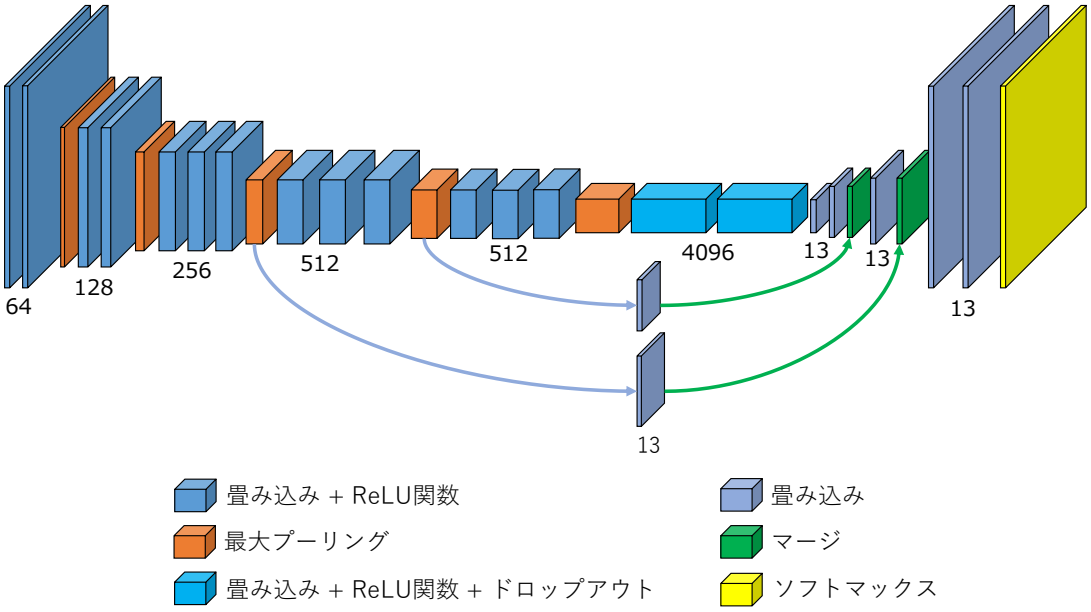


図 4: 提案手法のセマンティックセグメンテーションネットワーク．各ブロックはネットワークにおける層を表し，数字は各層の次元を表す．

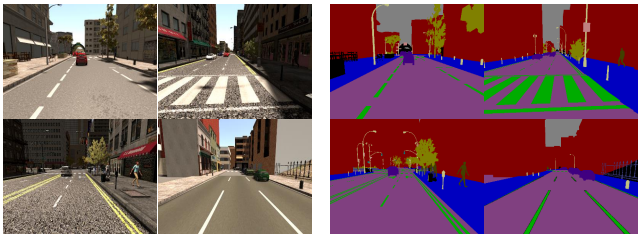
クセル一つに対して 61 次元の特徴量を持つ．従来手法 [7] と同様に，各スーパーピクセルの上下左右のスーパーピクセルの特徴も利用し，計 305 次元の特徴をラベル推定の入力とする．ラベル付き合成画像から上記のようにして得た特徴を用いて SVM を学習するが，学習時の正解ラベルは，合成画像の正解ラベルから各スーパーピクセルの範囲内において一番頻出なラベルとする．得られた特徴と正解ラベルを用いて，一つのスーパーピクセルに対して一つのラベルが出力される SVM が学習できる．合成画像と同様に実画像をスーパーピクセルに分割して特徴量を求めて，学習した SVM を使って実画像のラベルを推定する．しかし，ラベルはスーパーピクセル単位で推定しているため，画像中の小さい物体は同じスーパーピクセル内の周囲のピクセルの特徴の影響を受けて，他の大きな物体として推定される可能性が高い．そこで，SVM から予測の信頼度となる決定係数を得て，決定係数が大きい上位 60%のスーパーピクセルのみを，セマンティックセグメンテーション用のネットワークを学習する際に用いる．

4. 評価実験

提案手法の有効性について，合成画像のみを学習に用いた半教師あり学習を行わない手法や Zhang らの手法 [7] に対して，セマンティックセグメンテーションの識別精度を比較することで評価した．一方で，半教師あり学習を行わずにセマンティックセグメンテーションネットワーク自体の学習にオプティカルフローを利用する場合と提案手法を比較することで，実画像ラベルの推定を行う半教師あり学習にオプティカルフローを用いる必要性を確認した．また，提案手法における実画像ラベルの推定に，オプティカルフローの替わりに深度情報を用いた場合と提案手法を比較することで，実画像ラベルの推定におけるオプティカルフローと深度情報の有効性を比較した．

データセット

本研究では合成画像である SYNTHIA [5] データセットと実画像である Cityscapes [8] データセットを用いた．各データセットの画像例を図 5 と図 6 に示す．SYNTHIA は様々な環境の交通シーンを想定した合成画像のデータセットであり，RGB 画像，深度画像と物体毎にピクセル単位でラベルが振られたデータなどが含まれている．本研究では，車載カメラ映像を想定した時間的に連続な画像データセットである SYNTHIA VIDEO SEQUENCES の中で，進行方向を写したもので実験段階で利用可能である 1，2，4，5，6 シーンにおける Summer 環境下でのデータ計 4,535 枚の画像を用いた．物体のラベルとしては空，道路，車，歩行者など，合計 13 クラスがピクセル毎にラベル付けされている．Cityscapes は，実環境で車載カメラを用いて撮影された交通シーンの画像のデータセットであり，RGB 画像，視差画像と物体毎にピクセル単位でラベルが振られたデータなどが含まれている．物体のラベルとしては SYNTHIA の 12 クラスを含む合計 34 クラスが存在する．本研究では，連続画像でありピクセル毎にラベル付きのデータがある 5,000 枚の画像を用いたが，実際にはラベルデータを用いたのは評価のときだけであり，学習時にはラベルデータの真値は無いものとした．評価データには Cityscapes データセットにおける training セットから 500 枚を選択して用いた．



RGB画像　　真値ラベル画像

図 5: SYNTHIA [5] dataset の画像例

RGB画像　　真値ラベル画像

図 6: Cityscapes [8] dataset の画像例

実験方法

セマンティックセグメンテーション用のネットワークとして図 4 に示すネットワークを採用し，深層学習ライブラリとして Theano [12] をバックエンドに用いた Keras [13] を利用した．学習時の最適化関数は AdaDelta [14]，最適化のパラメータはデフォルト値を用いた．ミニバッチ中に含まれる画像数は合成画像から 4 枚，実画像から 4 枚の計 8 枚とした．合成画像のセマンティックセグメンテーションに関するロスはネットワークの出力結果と真値ラベルを用いて式 (1) を用いて計算し，実画像のセマンティックセグメンテーションに関するロスはネットワークの出力結果と 3.2 節述べた方法で推定したラベルを用いて同様に計算した．各データが 15 回以上使われるまでネットワークの重みを更新した．

評価尺度

セマンティックセグメンテーションの精度評価については，mean Intersection Over Union (mIOU) と呼ばれる評価手法を用いた．mIOU は，判定されるクラス数を n，正解データの i ラベルのピクセルの集合を A_i，推定結果の i ラベルのピクセルの集合を B_i とすると，

$$\mathrm{mIOU} = \frac{1}{n}\sum_{i=1}^{n}\frac{A_i \cap B_i}{A_i \cup B_i} \qquad (2)$$

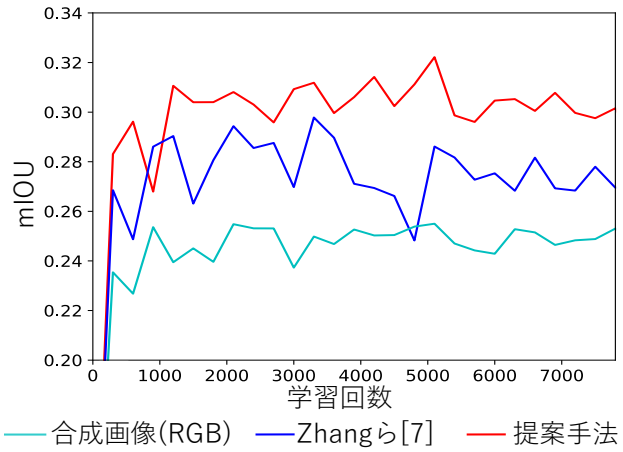


図 7: 合成画像のみで学習する手法，ラベル尤度のみで実画像ラベルの推定を行う手法，提案手法における mIOU 変化の比較

で表すスコアとなる．mIOU ではピクセル毎の一致率とは違い，誤った推定結果を出力するほどスコアが下がる．例えば一番出現頻度が高いラベルを優先して出力するモデルがあった場合，ピクセル毎の一致率は高くなる可能性はあるが，mIOU のスコアは下がることになる．また，出現頻度が低く，推定が難しいラベルに関して出力が上手くできいない場合，一致率ではほとんど影響は無いが，mIOU では全クラスの平均を取るのでスコアが大きく下がることになる．つまり，mIOU で評価することによって，モデルが複雑な出力を可能としているか評価できる．

4.1 提案手法の有効性

提案手法の有効性を示すために，提案手法に加えて以下 2 つの比較手法も実装，評価した．

合成画像のみ (RGB) 半教師あり学習を行わずに，ラベルデータが存在する合成画像のみで学習したネットワークを実画像に適用する手法である．

Zhang らの手法 [7] オプティカルフローを使わずに実画像ラベルを推定し，合成画像と実画像両方を用いて学習したネットワークを用いる．Zhang らの手法 [7] において，ラベル分布のロスは学習に加えない条件下で学習する手法である．

提案手法，合成画像のみ (RGB)，Zhang らの手法 [7] のネットワークを用いた評価データに対する mIOU の変化を図 7 に示す．また，それぞれのセマンティックセグメンテーションの結果例を図 8 に示す． 図 7 を見ると，合成画像のみ (RGB) の半教師あり学習を行わないネットワークよりも Zhang らの手法 [7] や提案手法のネットワークの方が mIOU の値が高いことから，簡単な方法にて実画像のラベルを推定し，その推定ラベルをネットワークの学習に

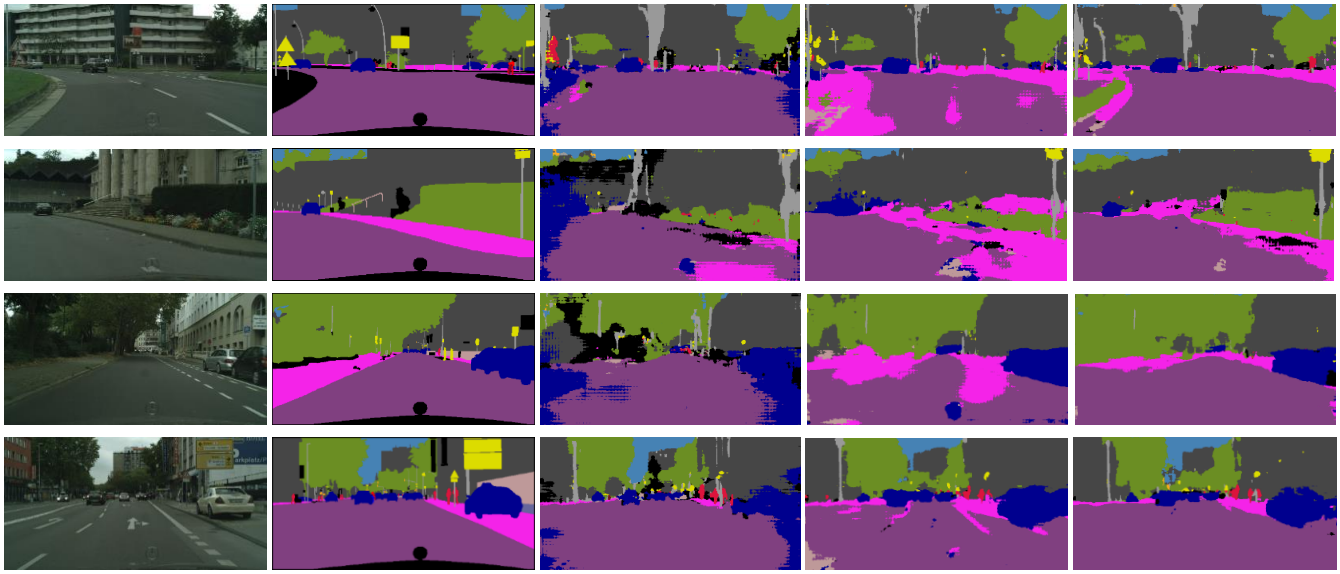


図 8: セマンティックセグメンテーションの結果例

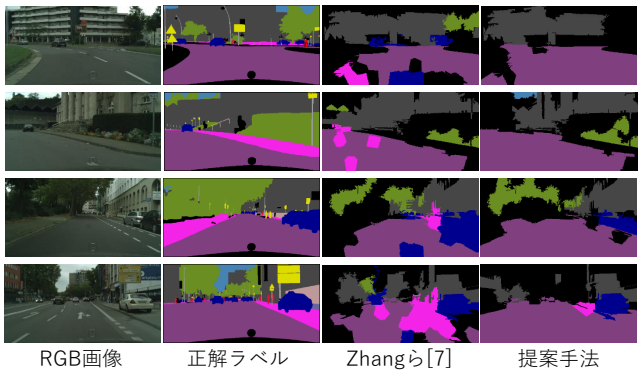


図 9: 実画像ラベルの推定例．黒くなっている部分は SVM の決定係数が小さく，セマンティックセグメンテーションネットワークの学習には使わない部分を表している．

用いる半教師あり学習手法の有効性を確認できる．また，Zhang らの手法 [7] よりも提案手法の方が優れていることから，ネットワークの学習に用いる実画像ラベル推定にオプティカルフロー特徴を追加することの有効性を確認できる．これは，ネットワークの学習方法は同じであるので，学習に用いる実画像ラベルの推定において提案手法が優れているといえる．

Zhang らの手法 [7] と提案手法におけるネットワークの学習に用いられる実画像のラベルの違いを図 9 に示す．図において黒くなっている部分は SVM の決定係数が小さく，セマンティックセグメンテーション用のネットワークの学習には使わない部分を表している．図 9 を見ると，Zhang らの手法 [7] においては隣り合う可能性が高い組み合わせである道路と歩道や道路と車の境界付近が上手く推定できていないことが分かる．Zhang らの手法 [7] においてはスーパーピクセル毎にラベルを推定する際，隣接スーパーピクセルの特徴量も推定に使われているので，隣接する物体の組み合わせを誤って推定することは少ない．しかし，周囲の特徴を考慮しても推定が難しくなる境界部分では誤って推定してしまうことが多くなると考えられる．一方で，提案手法での結果においては道路の部分が上手く推定できていることが分かる．推定時の特徴量としてオプティカルフローを追加したことで，推定スーパーピクセルの画像中でのおおよその位置情報が得られることによる道路と歩道の境界部分の推定精度の上昇や，動物体の検出により道路と車の境界部分の推定精度が向上することが確認できる．

次に，学習に用いる実画像ラベル推定の精度の向上が実際にセマンティックセグメンテーションの精度の向上に繋がっていることを各クラスの IOU を見ることで確認する．表 1 に Zhang らの手法 [7] と提案手法における各クラスの IOU を示す．表 1 から，道路，歩道と車のクラス IOU が Zhang らの手法 [7] よりも提案手法の方が大きく上昇していることから，学習に用いる実画像ラベル推定の精度向上の部分をセマンティックセグメンテーションネットワークが学習できていることが確認できる．

4.2 実画像ラベルの推定にオプティカルフローを用いる有効性

前節では，実画像ラベルの推定時にオプティカルフローを用いることが有効であることを確認した．ここで，セマンティックセグメンテーションネットワーク自体の学習にドメインに依らず似たものとなるオプティカルフローを用いることで，一度実画像ラベルの推定を行わずに，ドメイ

表 1: クラス IOU の比較

	Zhang ら [7]	提案手法
mean	0.281	0.301
Sky	0.569	0.567
Building	0.580	0.568
Road	0.475	0.663
Sidewalk	0.170	0.254
Fence	0.003	0.003
Vegetation	0.623	0.594
Pole	0.193	0.141
Car	0.339	0.443
Traffic Sign	0.087	0.059
Pedestrian	0.184	0.119
Bicycle	0.028	0.021
Traffic Light	0.031	0.035

ンの違いによる影響を受けにくいネットワークを学習できる可能性が考えられる．本節では次のように学習したネットワークと提案手法を比較する．

合成画像のみ (RGB+フロー) 提案手法とは異なるアプローチで，オプティカルフローをネットワークの学習に用いる手法の 1 つである．提案手法では，学習に用いる実画像ラベルの推定時にオプティカルフローを用いているが，この手法ではオプティカルフローを直接セマンティックセグメンテーション用のネットワークの学習に用いる．具体的には，図 4 において最初の畳み込み層の入力次元を 3 次元から 5 次元に変更し，RGB の 3 次元にオプティカルフローの 2 次元を追加したものを入力とするネットワークを構築する．また，合成画像の RGB 画像のみで学習する手法と同様に半教師あり学習は行わずに，ラベル付き合成画像のみを学習に用いて，実画像に対するセマンティックセグメンテーションの精度を評価する．

提案手法，合成画像のみ (RGB)，合成画像のみ (RGB+フロー) のネットワークを用いた評価データに対する mIOU の変化を図 10 に示す．図 10 より，合成画像のみ (RGB) と合成画像のみ (RGB+フロー) を比べると合成画像のみ (RGB+フロー) のネットワークの方がセマンティックセグメンテーションの精度が高いことから，オプティカルフロー特徴をネットワークの入力に加えることでも精度を向上させることが可能であると分かる．しかし，オプティカルフローの 2 次元だけを入力としたネットワークでは上手く学習できないことから，オプティカルフローはセマンティックセグメンテーション用のネットワークにおいて RGB 値の補助的な特徴でしかない．よって，合成画像のみで学習したネットワークでは，仮にオプティカルフローはドメイン共通であっても，RGB 値の部分においてドメインの違いの影響を受けてしまうと考えられる．そのため，部分的でも実画像での RGB 値に対応するラベルを学習し，

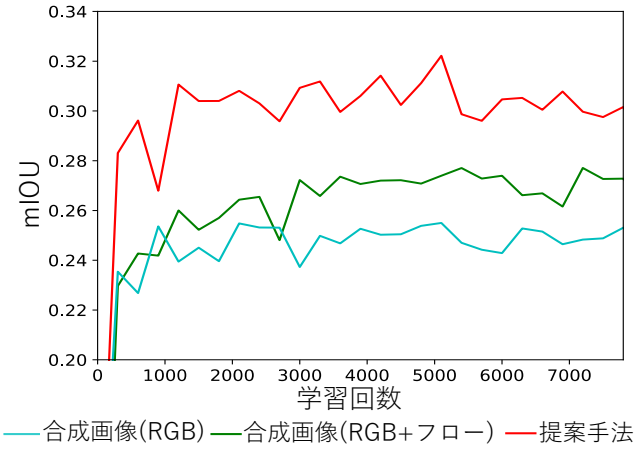


図 10: 合成画像の RGB 画像のみで学習する手法，セマンティックセグメンテーションネットワーク自体の入力にオプティカルフローを追加する手法，提案手法における mIOU 変化の比較

ドメイン適応することでセマンティックセグメンテーションの精度の向上に繋がったと考えられる．

4.3 深度情報を用いた実画像ラベル推定

前節では，実画像に対応するラベルを学習することの有効性が分かったが，そのラベル推定時に用いる特徴としてオプティカルフロー以外にも深度情報が考えられる．そこで，次のように学習したネットワークと提案手法を比較する．

提案手法 (深度) 学習に用いる実画像ラベルの推定時に，図 3 におけるオプティカルフローに替わって 1 次元の深度情報を用いて実画像ラベルの推定を行う手法である．車載カメラ映像であることにより似たような特徴量を持つのはオプティカルフローだけではなく，ピクセル単位でカメラからの物体までの距離を表す深度もその 1 つであるといえる．そこで提案手法のオプティカルフローと同様の手法にて，深度を用いて推定した実画像ラベルをセマンティックセグメンテーション用のネットワークの学習に用いる．

提案手法，Zhang らの手法 [7]，提案手法 (深度) のネットワークを用いた評価データに対する mIOU の変化を図 11 に示す．図 11 より，Zhang らの手法 [7] と提案手法 (深度) を比較すると提案手法 (深度) の方が優れていることから，深度が実画像ラベルの推定に有効な情報であることが確認できる．背景クラスとの深度の差から手前にある物体クラスの検出が可能となり，道路と車の境界部分の推定精度の向上に繋がっている．しかし，提案手法が提案手法 (深度) よりも優れていることから，ラベル推定時の特徴として深度情報よりもオプティカルフローが有効であることが分かる．深度情報は，多くの場面において距離がある程度一定

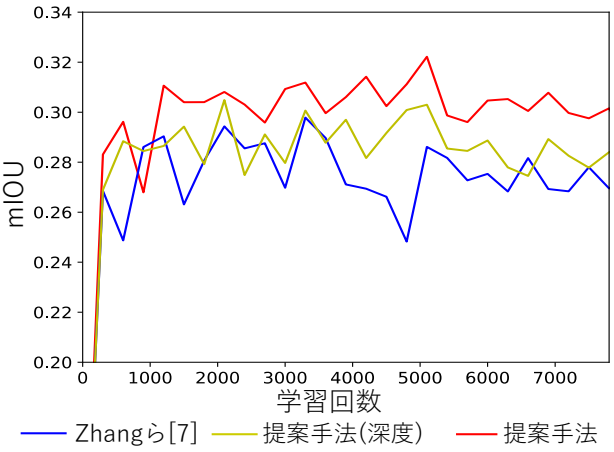


図 11: ラベル尤度のみで実画像ラベルの推定を行う手法，ラベル尤度に深度情報を加えて実画像ラベルの推定を行う手法，提案手法における mIOU 変化の比較

になるクラスに対しては有効な情報となるが，場面によって近くにあったり遠くにあったりするクラスにおいては，推定の助けとなる情報は得られない．また，オプティカルフローのように推定スーパーピクセルのおおよその位置情報は得られないことも，推定精度の差に生じていると考えられる．

5. 結論

本研究では，ラベル付き合成画像とラベル無し実画像を用いたセマンティックセグメンテーションを扱った．既存手法では，正解ラベルのない実画像に対して，一般物体識別用のデータセットでのラベル尤度から実画像のラベルを推定し，ネットワークの学習に用いる．そこで，車載カメラを想定した合成画像ならびに実画像で似た特徴量となるオプティカルフローを用いることで，実画像のラベル推定を改善し，セマンティックセグメンテーションネットワークの推定精度を向上させる手法を提案した．

提案手法の有効性を評価するため，合成画像として SYNTHIA[5] データセットを，実画像としては Cityscapes[8] データセットを用いて評価実験を行った．提案手法では，実画像ラベルの推定時にオプティカルフローを用いるだけで既存研究と比べてセマンティックセグメンテーションに対する精度の向上を確認した．加えて，オプティカルフローをそのままセマンティックセグメンテーションネットワークの入力として用いるよりも実画像の推定に用いる方が有効であること，実画像ラベルの推定において深度情報よりもオプティカルフロー情報の方が優れた特徴量であることを確認した．また，進行方向を写す車載カメラで撮られた映像であることに注目してオプティカルフロー特徴を用いているので，どちらも車載カメラで撮られた連続した画像データセットの組み合わせに対しては提案手法が有効ではないかと考えられる．

参考文献

[1] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, *Proceedings of Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012).
[2] Girshick, R., Donahue, J., Darrell, T. and Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587 (2014).
[3] Long, J., Shelhamer, E. and Darrell, T.: Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015).
[4] Everingham, M., Van Gool, L., Williams, C. K., Winn, J. and Zisserman, A.: The pascal visual object classes (voc) challenge, *Journal of computer vision*, Vol. 88, No. 2, pp. 303–338 (2010).
[5] Ros, G., Sellart, L., Materzynska, J., Vazquez, D. and Lopez, A. M.: 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).
[6] Hoffman, J., Wang, D., Yu, F. and Darrell, T.: Fcns in the wild: Pixel-level adversarial and constraint-based adaptation, *arXiv preprint arXiv:1612.02649* (2016).
[7] Zhang, Y., David, P. and Gong, B.: Curriculum domain adaptation for semantic segmentation of urban scenes, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2020–2030 (2017).
[8] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B.: The cityscapes dataset for semantic urban scene understanding, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3213–3223 (2016).
[9] Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A. and Brox, T.: FlowNet 2.0: Evolution of Optical Flow Estimation With Deep Networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2462–2470 (2017).
[10] Li, Z. and Chen, J.: Superpixel segmentation using linear spectral clustering, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1356–1363 (2015).
[11] Mottaghi, R., Chen, X., Liu, X., Cho, N.-G., Lee, S.-W., Fidler, S., Urtasun, R. and Yuille, A.: The role of context for object detection and semantic segmentation in the wild, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 891–898 (2014).
[12] Al-Rfou, R., Alain, G., Almahairi, A., Angermueller, C., Bahdanau, D., Ballas, N., Bastien, F., Bayer, J., Belikov, A., Belopolsky, A. et al.: Theano: A Python framework for fast computation of mathematical expressions, *arXiv preprint arXiv:1605.02688* (2016).
[13] Chollet, F. et al.: Keras: Deep learning library for theano and tensorflow (2015).
[14] Zeiler, M. D.: ADADELTA: an adaptive learning rate method, *arXiv preprint arXiv:1212.5701* (2012).