Vol.2017-CVIM-207 No.31
2017/5/10

適応的なドロップアウト空間の学習による セマンティックセグメンテーション

宮内 佑多朗[a)]　鮫島 正樹[b)]　菅野 裕介[c)]　松下 康之[d)]

概要：ニューラルネットワークを用いた画像認識において学習データに対する過学習を防ぐための手法の一つにドロップアウトがある．これはネットワーク中のノードをランダムに欠落させるものであるが，その最適なドロップアウト率はタスクごとに異なり，事前に調整する必要がある．さらに，セマンティックセグメンテーションのように空間的な特徴配置が重要な役割を果たすタスクにおいては，最適なドロップアウト率は領域毎に異なる可能性があるが，一様にドロップアウト率を決定する従来の枠組みではこのような最適化を行うことができない．本論文ではこれに対し，ドロップアウト率を空間的に最適化するための手法を提案する．従来手法の一様なドロップアウトに加え，提案手法では領域毎の最適なドロップ率を入力画像から適応的に決定するためのネットワークを追加し，セマンティックセグメンテーション精度が向上するようにネットワーク全体を学習する．複数の公開データセットを用いた評価実験により従来のドロップアウト手法との比較を行い，提案手法の有効性を示す．

Miyauchi Yutaro[a)]　Masaki Samejima[b)]　Yusuke Sugano[c)]　Yasuyuki Matsushita[d)]

1. 序論

ディープニューラルネットワーク (Deep Neural Network: DNN) は現在急速に普及・発展しつつあり，画像認識，音声認識などの幅広い分野で応用されている．特に画像認識の分野では，DNN の一種である畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) が様々なタスクに利用され，既存手法の性能を大きく向上させている [1].

CNN は学習用に与えられたデータに関しては非常に高い識別精度を実現出来る一方，学習データに対する過学習により未知のデータに対する精度が大きく下がる場合がある．この過学習を抑制するために，入力を圧縮することでパラメータの数を削減するプーリング [2] に加え，ネットワーク中のノードをランダムに欠落させるドロップアウト [3] がしばしば用いられる．しかし，ドロップアウトにおいてノードを欠落させる確率はモデル決定時のハイパーパラメータとなり，適切なドロップアウト率は手動で調整する必要がある．正解データと出力の誤差が小さくなるようなドロップアウトを学習させ，入力からドロップアウト率を決定する研究もなされているが [4]，教師とするドロップアウト率を得るため，出力の数に比例した回数分誤差を計算する必要があり，計算量が大きくなるという問題がある．

また，ドロップアウトでは，欠落させる確率は入力画像の全領域について一定となる．画像認識のタスクによっては，入力画像の各領域でドロップアウト率を変化させ，空間的にノードを欠落させた方が良い場合があると考えられる．例えば，画像中の物体を画素レベルで識別・分割するセマンティックセグメンテーション [5], [6], [7] において，CNN は入力画像の各画素のラベルを個別に推定する必要がある．この場合，画像の各チャンネルの領域を空間的にドロップアウトさせることによって領域毎のクラス分類の精度を向上させることが出来ると考えられる．

本論文では，入力画像をもとに適応的にドロップアウト率を決定させることによって，空間的にドロップアウトを行う手法を提案する．通常のドロップアウトのように空間的に一定の確率によってドロップアウトを決定するのではなく，提案手法は学習結果によりその欠落させる確率を変化させる．さらに，提案する手法は空間的なドロップアウト率を推定するための層を CNN に追加し，入力画像に応じてドロップアウト率を変化させるような構造を学習す

a) 大阪大学院情報科学研究科，miyauchi.yutaro@ist.osaka-u.ac.jp
b) 大阪大学院情報科学研究科，samejima@ist.osaka-u.ac.jp
c) 大阪大学院情報科学研究科，sugano@ist.osaka-u.ac.jp
d) 大阪大学院情報科学研究科，yasumat@ist.osaka-u.ac.jp

る．複数のデータセットを用いた評価実験により，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寄与することを示す．

2. 関連研究

ドロップアウト率の最適化

先述の通り，ドロップアウト率を学習の中で最適化する方法は過去にいくつか提案されている．Adaptive Dropout は，最初にドロップアウトが適用される層の各出力それぞれにドロップアウトの正解データを用意し，次にその正解に近くなるようなドロップアウト率を出力するようにパラメータの更新を行う手法である [4]．ここでの正解データは，ある出力に関して，ドロップアウトした場合とドロップアウトしない場合の損失の比較を行い，損失の少ない方を正解としたものであるため，ドロップアウト対象の要素が多い場合の計算量が問題となる．また，ベイズ推論を用いて最適なドロップアウト率を推定する手法も提案されている [8]．ただし，この手法ではモデルが収束するまでに通常よりもかなりの学習回数，つまり計算量を必要としている．提案手法はこれらの手法とは異なり，ドロップアウト率の学習に誤差逆伝搬法を用いることで計算量の削減を行うほか，ドロップアウト率の空間的な最適化を行う．

CNN の空間的な最適化

画像の空間性を考慮して CNN の構造を最適化する手法も盛んに研究が行われており，その一つに Spatial Transformer Networks [9] がある．CNN の中間に挟まれた Spatial Transformer 層によって，識別する対象のデータがある特徴量マップ内の空間を抽出し，入力画像中に含まれる識別対象のスケール，傾き，歪みなどの多様性に影響されないクラス分類の CNN を構築する手法を提案している．また，Pooling を繰り返し適用することで失われる画像の空間的情報の正確さを，Pooling 適用前の特徴量マップを再利用することで復元し，ポーズ推定に必要な関節の位置推定でより高い性能を発揮する CNN モデルも提案されている [10]．提案手法の目的も同様に，セマンティックセグメンテーションのタスクにおける画像特徴の空間性を考慮することであるが，ドロップアウト率を空間的に最適化することによって CNN の性能を向上させる点でこれらの研究と異なる．

3. 提案手法

提案手法の概要を図 1 に示す．提案手法のモデルは Convolutional Encoder-Decoder (CED) [11] の構造を元にしている．CED では，まずネットワークの前半部分の Encoder で入力画像の特徴を抽出しつつ，繰り返し Pooling を行うことで中間層のノードを減らし次元削減する．そして，Encoder で抽出した特徴量マップをネットワーク後半の Decoder で元の画像サイズに戻す．セマンティックセグメンテーションのタスクでは，図 1 に示すような各画素に対し正解ラベルが与えられているラベル画像を教師とし，損失関数には Softmax Cross Entropy を用いる [5]．正解クラスの総数を n，正解データ $\boldsymbol{t} \in \mathbb{R}^n$ の要素を $t_i \in \{0, 1\}$，ネットワーク出力 $\boldsymbol{y} \in \mathbb{R}^n$ の要素を y_i とすると，Softmax Cross Entropy 関数 E は

$$E(\boldsymbol{t}, \boldsymbol{y}) = -\frac{1}{n}\sum_{i=1}^{n}\{t_i \log y_i + (1 - t_i)\log(1 - y_i)\} \quad (1)$$

と定義される．

CED では過学習を防ぐため，中間層で抽出された特徴量マップに対して，全領域で一様のドロップアウト率による通常のドロップアウトを適応している．提案手法では通常のドロップアウトに加え，空間的にドロップアウト率を決定するための畳み込み層 (以下，確率決定層) を用意し，確率決定層の出力をもとに通常の畳み込み層 (以下，通常層) のノードを欠落させる．確率決定層は通常層と同じ構造をしており，通常層と同じ数の出力を持つ．確率決定層での出力はすべて 0 から 1 の間の数値で出力され，この数値が高いほど，対応する通常層の出力がそのまま出力される確率が高くなる．確率決定層の確率でドロップアウトを適用することで空間的なドロップアウトを行うことができので，これを空間適応型ドロップアウトとする．通常層のパラメータの学習と並行して確率決定層のパラメータも学習を行うことで，入力画像に対して適応的かつ空間的にドロップアウト率を変更する CNN を実現している．

3.1 確率決定層

確率決定層の詳細を図 2 に示す．確率決定層のノードの出力は，対応する通常層のノードを欠落させるための確率として扱える形である必要があるため，活性化関数としてシグモイド関数を使用する．通常層の出力 $\boldsymbol{h} \in \mathbb{R}^n$ の要素を h_i，確率決定層の出力 $\boldsymbol{r} \in \mathbb{R}^n$ の要素を r_i とすると，空間適応型ドロップアウト層の最終的な出力 $\boldsymbol{y} \in \mathbb{R}^n$ は

$$\boldsymbol{y} = \boldsymbol{h} \circ \boldsymbol{m} = \begin{pmatrix} h_1 m_1 \\ h_2 m_2 \\ \vdots \end{pmatrix} \quad (2)$$

$$P(m_i = 1) = r_i$$

$$P(m_i = 0) = 1 - r_i$$

となる．すなわち，$\boldsymbol{m}$ の要素 m_i は対応する $\boldsymbol{r}$ の要素 r_i の確率によって 0 か 1 の値を取り，$\boldsymbol{r}$ はノード毎のドロップアウト率を表現することになる．

確率決定層のパラメータの更新には通常のパラメータの学習と同じく誤差逆伝搬法を用い，通常層の学習と確率決定層のパラメータの学習を同時に行う．ただし，分類問題であるセマンティックセグメンテーションの損失関数として Softmax Cross Entropy を利用する一方，確率決定層の

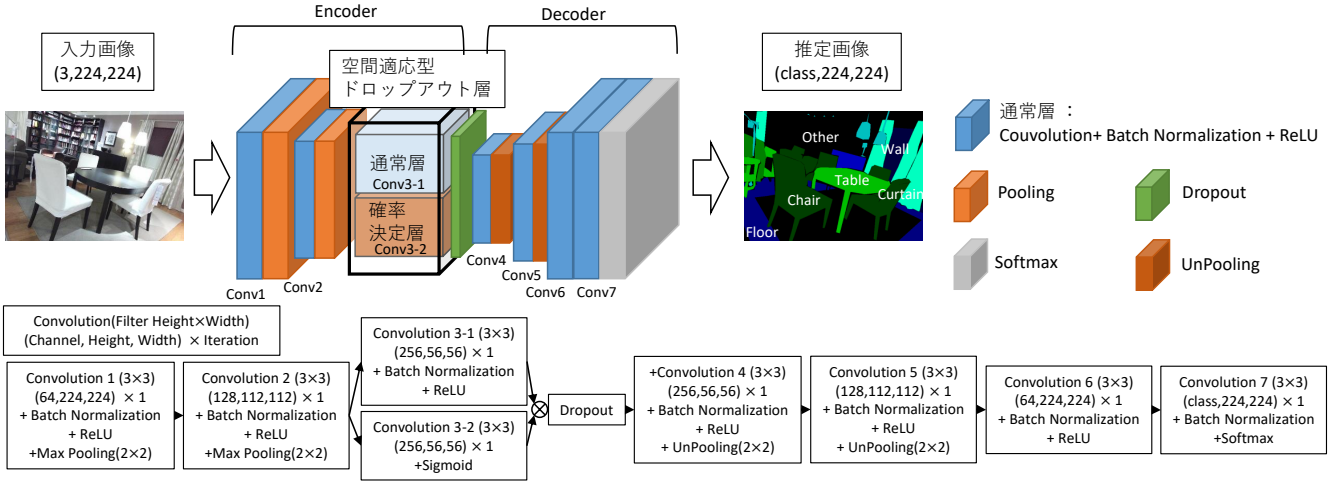


図 1: 提案手法の概要．Convolutional Encoder-Decoder ネットワークにおいて，Encoder の最後に空間的に異なるドロップアウト率を出力するための畳み込み層を伴う空間適応型ドロップアウト層を挿入する．その後更に通常のドロップアウト層を適用し，Docoder のネットワークにより各画素のクラスを推定する．ネットワーク構造の詳細を図の下部に示す．

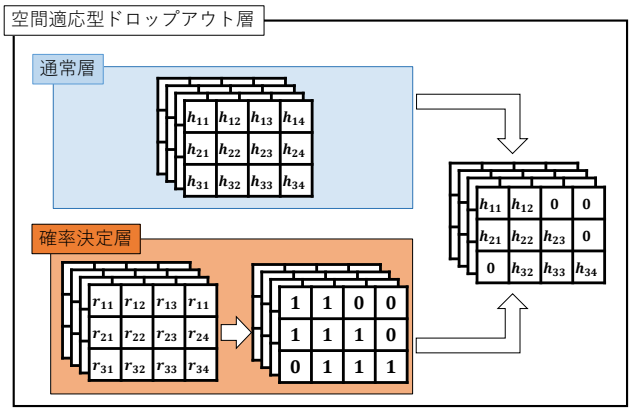


図 2: 確率決定層の模式図．確率決定層から出力される確率 $\boldsymbol{r}$ に応じて二値のマスク $\boldsymbol{m}$ が生成され，これによって通常層の出力 $\boldsymbol{h}$ のどの要素を欠落させるかが決まる．

出力であるドロップアウト率は連続値であるため，まず更新ごとに目標とするドロップアウト率を計算し，目的値との二乗誤差を損失関数として学習を行う．

更新前のドロップアウト率 $\boldsymbol{r}$ の要素を r_i，ネットワーク全体の損失関数を E，ドロップアウト率 r_i に対応した通常層の出力 $\boldsymbol{h}$ の要素 h_i の伝搬誤差 $\frac{\partial E(\boldsymbol{t},\boldsymbol{y})}{\partial h_i}$ を δ_i，層全体の伝搬誤差の平均を $\overline{\delta}$，更新前のドロップアウト率 $\boldsymbol{r}$ の平均を $\overline{r}$ とすると，目標とするドロップアウト率 $\hat{\boldsymbol{r}}$ の要素 $\hat{r}_i$ は次のように定義される．

$$\hat{r}_i \leftarrow r_i - \gamma(r_i)(\delta_i - \overline{\delta}) \tag{3}$$

$$\gamma(r_i) = \frac{\alpha}{\beta(\overline{r} - r_i)^2 + 1} \tag{4}$$

式 (3) において学習率 γ は r_i の関数となっており，ドロップアウト率 r_i がドロップアウト率の平均 $\overline{r}$ から離れるほど学習率が小さくなる．また，式 (4) での α は全体の学習速度を調整し，β は更新幅の逓減率を調整する．

誤差逆伝搬法を用いた通常層のパラメータの更新において，式 (3) の伝搬誤差 δ_i は，値が正であるとき通常層のパラメータは負の方向に更新が進み，そのパラメータが属するノードの出力 h_i は小さくなる．逆に δ_i の値が負であるときにはパラメータは正の方向に更新が進み，そのパラメータが属するノードの出力 h_i も大きくなる．ここで式 (3) の式により，あるノードのドロップアウト率 r_i は，伝搬誤差 δ_i が負，つまり対応する出力 h_i が大きくなるときには，そのドロップアウト率 r_i も大きくなる方向に更新が進むことになる．逆に出力 h_i が小さく，0 に近づくときにはその出力のドロップアウト率 r_i も小さくなる．以上により，大きい出力をするノードは出力される確率が高く，小さい出力をするノードはドロップアウトする確率の高くなる確率決定層の出力値 $\hat{\boldsymbol{r}}$ が得られる．また，式 (3) では，偏差をとることによって全体のドロップアウト率が一定の方向に偏ることを防ぐ．

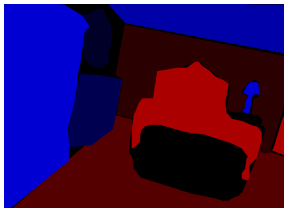
次に，確率決定層のパラメータ $\boldsymbol{g}$ をこのドロップアウト率を出力するように更新をする．ここで，パラメータの更新のための新たな目的関数 E_r を $\hat{r}_i, r_i$ の二乗誤差関数とすると，更新式は

$$\boldsymbol{g} \leftarrow \boldsymbol{g} - \frac{\partial E_r(\hat{r}_i, r_i)}{\partial \boldsymbol{g}} \tag{5}$$

と定義される．式 (3) から更新式は

$$\begin{aligned}
\frac{\partial \sigma(x)}{\partial x} &= \sigma(x)(1-\sigma(x)) \\
\boldsymbol{g} &\leftarrow \boldsymbol{g} - \gamma(r_i)\frac{\partial E_r(\hat{r}_i, r_i)}{\partial r_i}\frac{\partial r_i}{\partial \sigma(\boldsymbol{g}^T\boldsymbol{x})}\frac{\partial \sigma(\boldsymbol{g}^T\boldsymbol{x})}{\partial \boldsymbol{g}} \\
&= \boldsymbol{g} - \boldsymbol{x}\gamma(r_i)(r_i - \hat{r}_i)\sigma(\boldsymbol{g}^T\boldsymbol{x})(1-\sigma(\boldsymbol{g}^T\boldsymbol{x})) \\
&= \boldsymbol{g} + \boldsymbol{x}\gamma(r_i)(\delta - \overline{\delta})\sigma(\boldsymbol{g}^T\boldsymbol{x})(1-\sigma(\boldsymbol{g}^T\boldsymbol{x}))
\end{aligned} \tag{6}$$

のように変形される．σ は活性化関数として用いるシグモ

RGB画像　　正解ラベル画像

図 3: SUNRGBD 画像例 [13]


RGB画像　　正解ラベル画像

図 4: Stanford Background Dataset の画像例 [14]

イド関数である．

4. 評価実験

提案手法の有効性を示すために，セマンティックセグメンテーションのデータセットを用いて評価実験を行った．提案手法のほか，
(1) 通常の畳み込み層のみのモデル
(2) 通常のドロップアウトのみを適応したモデル
(3) 空間適応型ドロップアウトのみを適応したモデル
と性能の比較を行う．各モデルの基本構造は図 1 に準じ，提案手法における空間適応型ドロップアウト層と通常ドロップアウト層の組み合わせがそれぞれ (1) ドロップアウトのない通常層のみ，(2) 通常ドロップアウト層のみ，(3) 空間適応型ドロップアウト層のみに置き換わったモデルとなる．ドロップアウト率は，以後特に表記のない場合 0.5 とする．ネットワークパラメータ更新の最適化手法としては Adam [12] を使用した．Adam は確率決定層のパラメータ更新にも適用されている．また今回の実験において，学習率の式 (4) におけるハイパーパラメータは $\alpha = 10^{-7}, \beta = 10^2$ として実験を行った．

評価用のデータセットとしては，SUNRGBD [13]，Stanford Background Dataset [14] の二つを使用した．SUNRGBD は深度カメラを使って得られた画像のデータセットであり，RGB 画像，深度画像に加え，セマンティックセグメンテーションの真値データ，シーンの正解ラベルなどが，約 10,000 セット集められている．画像は全て室内で撮影されたもので，セマンティックセグメンテーションのラベルとして，机，椅子，壁，床，天井，本など計 38 クラスが画素毎にラベル付けされている (図 3)．今回の評価実験では，RGB 画像と，セマンティックセグメンテーションのラベル画像のみを用いた．学習には 5,000 枚の画像を使用し，1,000 枚を評価用のデータとした．

また，Stanford Background Dataset は主に屋外で撮影された画像が集められたデータセットで，RGB 画像と，セマンティックセグメンテーションの真値データがが 715 セット用意されている．セグメンテーションのラベルとしては空，木，道路，水など，合計 9 クラスが定義されている (図 4)．学習には 600 枚，評価には 100 枚の画像を使用した．

評価尺度としては，画素一致率と Mean Intersection Over Union (mIOU) [6] を用いた．判定されるクラス数を n，正解画像の i ラベルの画素の集合を A_i，推定画像の i ラベルの画素の集合を B_i，正解画像・推定画像の画素の総数を N すると，画素一致率は

$$\text{画素一致率} = \frac{\sum_{i=1}^{n} A_i \cap B_i}{N} \tag{7}$$

と定義される．ここでの積集合は，正解画像と推定画像の対応する画素がどちらも同一のクラスである画素とする．

また，mIOU は各クラスの Intersection Over Union (IOU) を平均したものとして

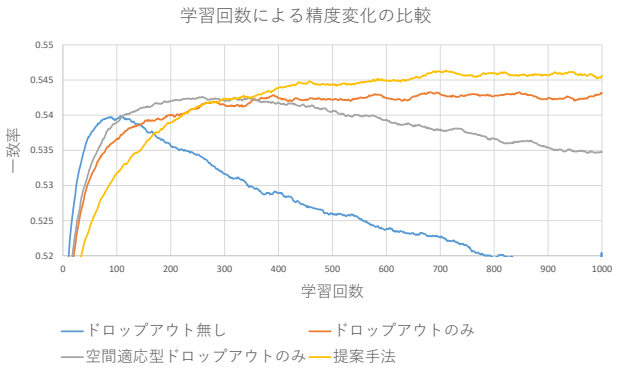
$$\text{mIOU} = \frac{1}{n} \sum_{i=1}^{n} \frac{A_i \cap B_i}{A_i \cup B_i} \tag{8}$$

と定義される．分母に和集合をとるため，誤った推定結果を出力するほど mIOU スコアが低下する．また，mIOU では全クラスの平均を取るため，出現数が少ないラベルに関して出力ができていない場合もスコアが大きく下がることになる．

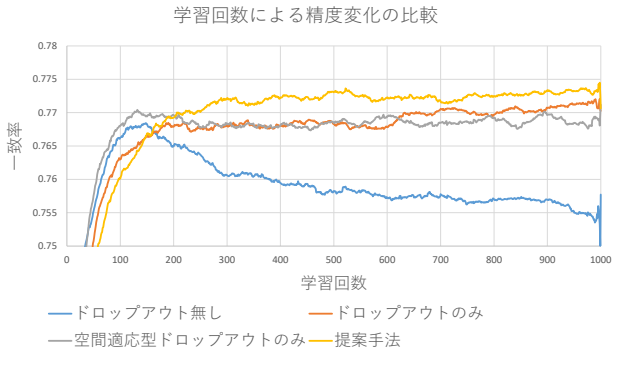
4.1 セグメンテーション性能の比較

二つのデータセットにおいて，画素一致率により各モデルの精度比較を行った結果を図 5 に示す．図 8a が SUNRGBD，図 8b が Stanford Background Dataset に対応しており，縦軸は画素一致率 (%)，横軸は学習回数で，各グラフはそれぞれ上述の 4 つのモデルの学習回数による画素一致率の変化を表している．図の可読性を考慮し，グラフを周辺 50 個の移動平均線とした．また，同様に mIOU の比較を行った結果を図 6 に示す．図 6a が SUNRGBD，6b が Stanford Background Dataset に対応し，縦軸が mIOU スコアを示す．

これら全ての比較において，ドロップアウトを一切含まない通常の畳み込み層のみのモデルが最も低い性能を示している．これらのデータセット，タスクにおいてはドロップアウトがモデルの性能に大きく影響していることがわかる．しかし，図 5 では，どちらのデータセットにおいても空間適応型ドロップアウトのみでは過学習を完全に抑制できていないことがわかる．空間適応型ドロップアウト層は

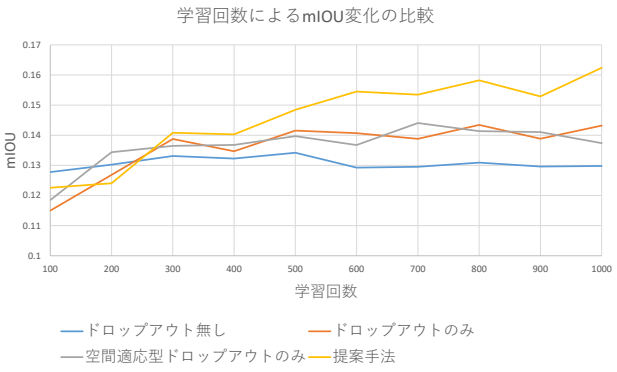


(a) SUNRGBD

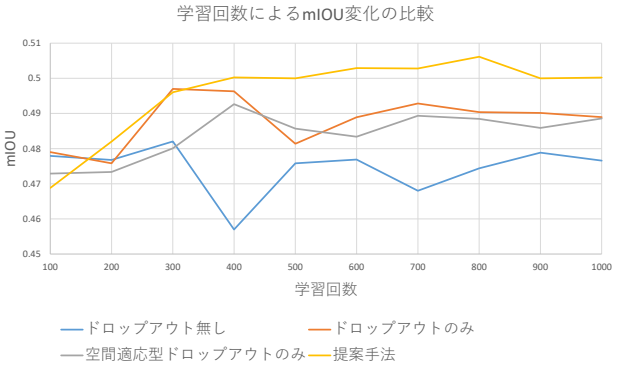


(b) Stanford Background dataset

図 5: 画素一致率の比較．縦軸は画素一致率 (%)，横軸は学習回数で，各グラフはそれぞれ上述の 4 つのモデルの学習回数による画素一致率の変化を表している．



(a) SUNRGBD



(b) Stanford Background dataset

図 6: mIOU の比較. 縦軸は mIOU，横軸は学習回数で，各グラフはそれぞれ上述の 4 つのモデルの学習回数による mIOU の変化を表している．

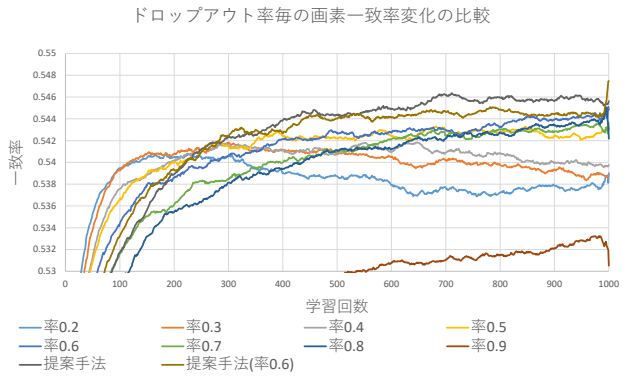
ドロップアウト率を学習することで性能の向上が期待できる一方，完全にランダムなドロップアウトを行うことによる過学習抑制効果は薄れていることがわかる．提案手法は通常ドロップアウトをさらに組み合わせることでこの欠点を補い，画素一致率において最も高い性能を実現している．一方，図 6 に示すように，mIOU に関しては空間適応型ドロップアウトのみのモデルでも通常のドロップアウトのみのモデルに比べ精度が落ちず，提案手法が最も良い結果を示していた．

さらに，提案手法には全体的なドロップアウト率を最適化する効果もあるため，空間適応型ドロップアウト層の効果を厳密に検証するために，通常ドロップアウト層のみのモデルのドロップアウト率を変化させて性能の比較を行った．図 7 は提案手法に加え，通常ドロップアウト層のみのモデルにおいてドロップアウト率を 0.1 刻みに変更して学習を行った際の画素一致率の変化を示している．さらに，提案手法のネットワークにおいても各データセットで最も性能の高いドロップアウト率を通常ドロップアウト層に採用した場合の性能も追加で示した．同様に，図 8 では上記の中から各データセットにおいて性能が上位 2 つのドロップアウト率を選択し，提案手法および最適なドロップアウト率の提案手法と合わせて各モデルの mIOU を示している．Stanford Background Dataset の画素一致率ではほぼ同等の結果が得られているものの，それ以外の場合では通常のドロップアウトと空間適応型ドロップアウト層を組み合わせる手法が最も高い性能を示すことが確認できた．提案手法の効果は全体的なドロップアウト率を最適化するだけに留まらず，空間的な最適化による性能向上が得られていることが確認できる．

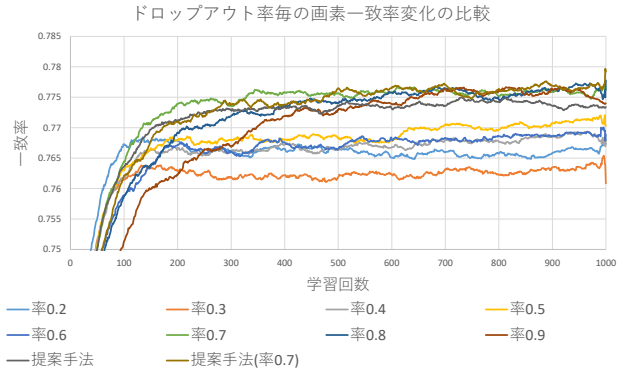
4.2 適応的ドロップアウト構造による注目度の変化

今回の評価実験で使用したモデルのうち，通常のドロップアウトのみ適応したモデルと提案手法のモデルそれぞれの推定結果画像を図 9 に示す．通常のドロップアウトのみのモデルに比べ，提案手法を用いることでクラス領域中にあらわれるノイズが減少する傾向が見られた．しかし，中には図 9 の 4 行目のように，提案手法の方がノイズを含む結果となっている場合も見られる．

図 10 ではこれら二つのモデルの違いをより詳細に可視化するため，各モデルがテスト時に注目する領域の比較を行った [15]．図 10 で示されている注目度マップは，その画素周辺一定領域を隠すことによってモデルの出力する推定画像の IOU がどれだけ悪化するかをプロットしたもの

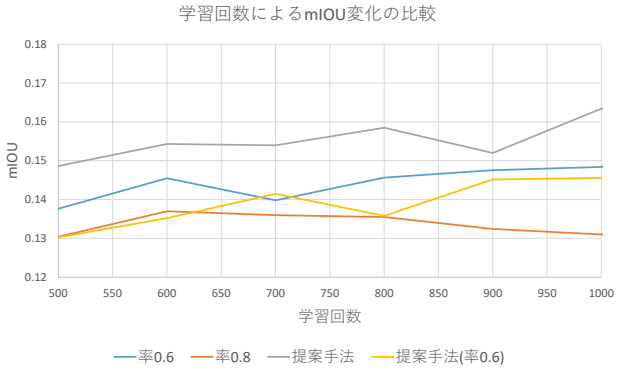


(a) SUNRGBD

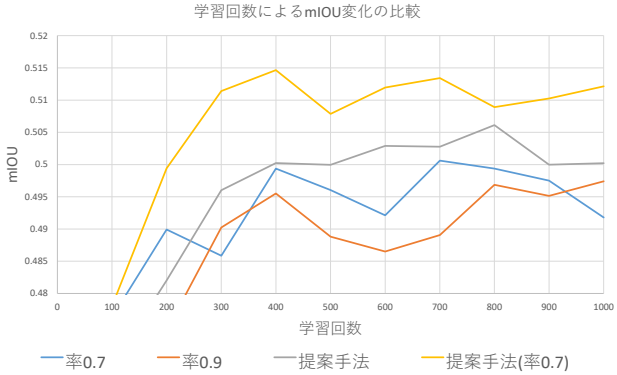


(b) Stanford Background dataset

図 7: ドロップアウト率毎の画素一致率．横軸は学習回数で，各グラフは学習回数による画素一致率の変化を表している．図中の率 0.*で表されているグラフは，通常のドロップアウトを適応したモデルのドロップアウト率を変更したモデルの実験結果を示す．



(a) SUNRGBD



(b) Stanford Background dataset

図 8: ドロップアウト率毎の mIOU 比較．横軸は学習回数で，各グラフは学習回数による mIOU の変化を表している．図中の率 0.*で表されているグラフは，通常のドロップアウトを適応したモデルのドロップアウト率を変更したモデルの実験結果を示す．

となる．具体的には，224 × 224 画素の元画像の各領域にサイズが 28 × 28 で要素が全て 0 のマスクをスライド幅 7 画素でかけた画像を生成し，マスク画像に対して CNN を適用して得られる IOU と，元画像の IOU との差を注目度としている．図 10 の注目度マップは，こうして得られる 49 × 49 画素の注目度マップを平滑化し，元画像のアスペクト比で表示したものとなる．提案手法では，図 10a の上画像のようにあるクラスを出力する上で正解領域の注目度を向上させるほか，図 10 下の画像のように正解領域以外の注目度を低下させる効果もあり，これらの組み合わせにより複合的にセマンティックセグメンテーションの性能を向上させていると言える．しかし，図 10b で示すように一部のクラスのセマンティックセグメンテーションでは正解領域周辺の情報によって IOU が向上する場合もある．

5. 結論

本論文では，学習によって適応的にドロップアウト率を変更する CNN によるセマンティックセグメンテーション手法を提案した．提案手法では一定の確率で CNN ノード

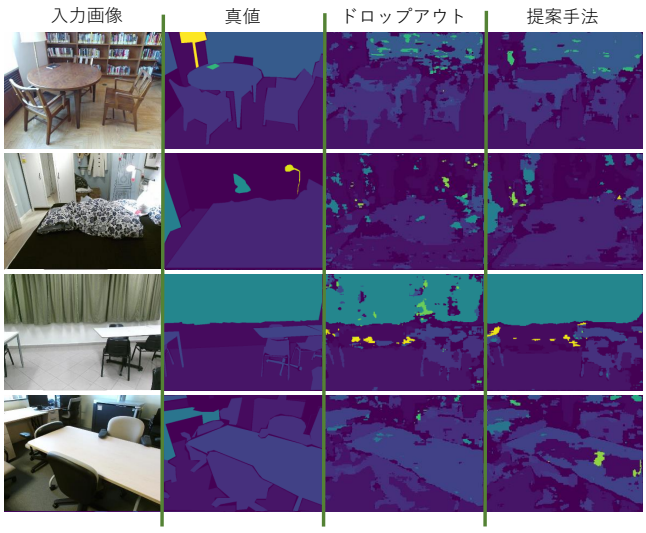
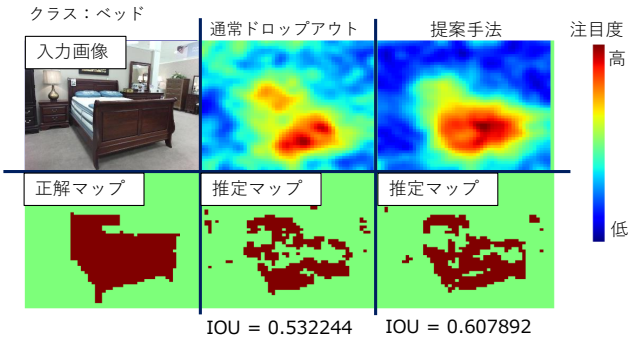
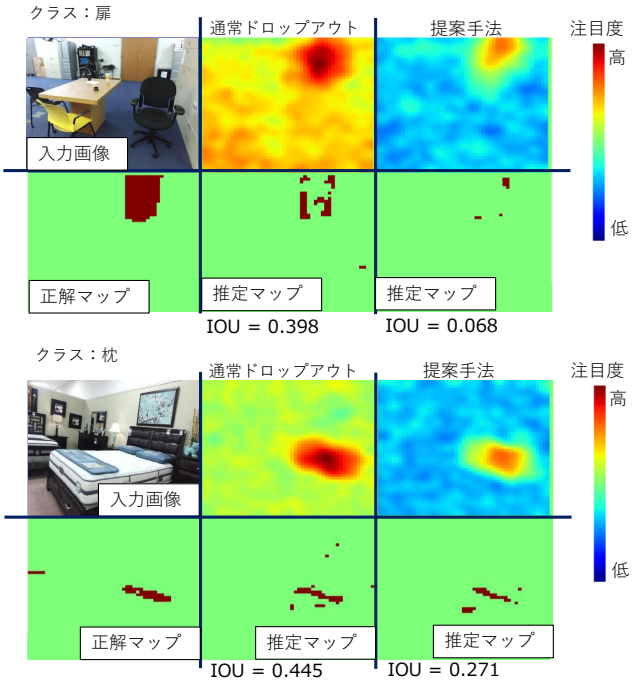


図 9: セマンティックセグメンテーションの推定結果の例．各列は左から入力画像，真値画像，通常のドロップアウトのみ適応したモデルの推定画像，空間適応型ドロップアウトのモデルの推定画像となる．



(a) 提案手法が有効であった例



(b) 提案手法が悪影響を及ぼした例

図 10: CNN の注目度マップの例．一行目は左から順に入力画像，通常のドロップアウトのみ適応したモデルの注目度マップ，空間適応型ドロップアウトのモデルの注目度マップに対応し，二行目は左から順にあるクラスの正解マップ，通常のドロップアウトを適応したモデルによる推定マップ，空間適応型ドロップアウトのモデルによる推定マップとなる．右に示した凡例のように，注目マップが赤いほど注目度が高く，各モデルが出力のためにその領域の情報を注目していることを表す．

の欠落を行う既存のドロップアウト層に加え，通常の畳み込み層に加えて学習された確率決定層の出力に応じて，画像領域ごとに異なる率でドロップアウトを行う．確率決定層の学習にも誤差逆伝搬法を利用することで，通常の畳み込み層と同時にネットワーク全体を学習することが可能となる．

SUNRGBD, Stanford Background Dataset の二つのデータセットを用いた評価実験では，提案手法を用いたモデルの mIOU スコアがベースライン手法から向上しており，空間適応型ドロップアウトと通常ドロップアウトを組み合わせる提案手法の有効性が確認できた．空間適応型ドロップアウトと通常ドロップアウトは，それぞれ CNN の空間構造の最適化と過学習防止という異なる目的・性質を持っており，セマンティックセグメンテーションにおいてはこれらの組み合わせが性能向上に大きく寄与していると考えられる．

今後の課題としては，今回実験で用いたネットワーク構造以外の場合でも提案手法の空間適応型ドロップアウトがセマンティックセグメンテーションの精度向上に有効か確認するほか，画像分類問題など他のタスクにおいても提案手法のアプローチが有効か検証することが挙げられる．

参考文献

[1] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp. 1097–1105 (2012).
[2] Lee, H., Grosse, R., Ranganath, R. and Ng, A. Y.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, *Proceedings of the 26th annual international conference on machine learning (ICML-9)*, pp. 609–616 (2009).
[3] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting, *Journal of Machine Learning Research*, Vol. 15, No. 1 (2014).
[4] Ba, J. and Frey, B.: Adaptive dropout for training deep neural networks, *Advances in Neural Information Processing Systems*, pp. 3084–3092 (2013).
[5] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A. L.: Semantic image segmentation with deep convolutional nets and fully connected crfs, *arXiv preprint arXiv:1412.7062* (2014).
[6] 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).
[7] Noh, H., Hong, S. and Han, B.: Learning deconvolution network for semantic segmentation, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1520–1528 (2015).
[8] Maeda, S.-i.: A bayesian encourages dropout, *arXiv preprint arXiv:1412.7003* (2014).
[9] Jaderberg, M., Simonyan, K., Zisserman, A. et al.: Spatial transformer networks, *Advances in Neural Information Processing Systems*, pp. 2017–2025 (2015).

[10] Tompson, J., Goroshin, R., Jain, A., LeCun, Y. and Bregler, C.: Efficient object localization using convolutional networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 648–656 (2015).
[11] Badrinarayanan, V., Kendall, A. and Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *arXiv preprint arXiv:1511.00561* (2015).
[12] Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
[13] Song, S., Lichtenberg, S. P. and Xiao, J.: SUN RGB-D: A RGB-D scene understanding benchmark suite, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 567–576 (2015).
[14] Gould, S., Fulton, R. and Koller, D.: Decomposing a scene into geometric and semantically consistent regions, *Proceedings of IEEE 12th International Conference on Computer Vision*, pp. 1–8 (2009).
[15] Zeiler, M. D. and Fergus, R.: Visualizing and understanding convolutional networks, *Proceedings of European Conference on Computer Vision*, pp. 818–833 (2014).