

1. はじめに

変化点検出とは、同一地点を異なる時刻に撮影した画像から変化した領域を抽出する問題であり、デジタル地図の自動更新や災害時の被害状況把握への活用が期待されている。従来の変化点検出 [2] は変化した領域のみを検出するため、変化領域が何であるかは不明である。そこで本研究では、U-Net[1] をマルチタスク化し、セマンティックセグメンテーションと変化点検出を同時に行う手法を提案する。また、各タスクに適切な学習重みを選択し精度向上を図る。

2. U-Net

U-Net は、Encoder-Decoder 構造のセマンティックセグメンテーション手法の 1 つである。U-Net のネットワーク構造を図 1 に示す。Encoder 部分と Decoder 部分の対応した解像度の特徴マップを連結させることで、大域的な特徴と局所的な特徴を捉えることができる。

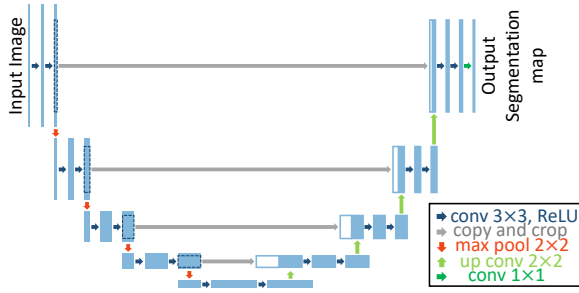


図 1: U-Net のネットワーク構造

3. 提案手法

本研究では、1 つのネットワークでセマンティックセグメンテーションと変化点検出を行う、U-Net を用いたマルチタスクラーニングを提案する。

3.1. U-Net を用いたマルチタスクラーニング

U-Net を用いたマルチタスクラーニングは、U-Net の Decoder 部分を、図 2 に示すようにタスクごとに分割したネットワークである。これにより、複数のタスクを同時に学習することができるため、ネットワークの数を減らすことができる。

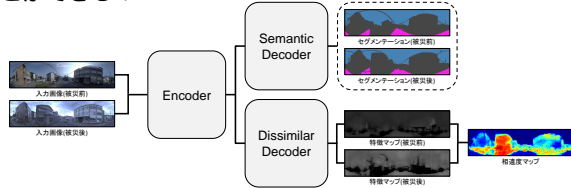


図 2: U-Net を用いたマルチタスクラーニングの流れ

3.2. 損失の計算

学習に使用する損失関数は、タスクごとに出力された損失を重み付き和で算出して使用する。セグメンテーションタスクの損失 各クラスに対する確率を特徴マップよりソフトマックス関数 \hat{y} を用いて算出する。次に、算出した確率を使用し、セグメンテーションの損失 L_{Seg} を求める。 L_{Seg} の算出式を式 (1) で表す。この時、 c はクラス、 h, w は入力画像の縦と横の幅である。

$$L_{Seg} = - \sum_{c=0}^C \sum_{j=0}^w y_j^c \log \hat{y}_j^c \quad (1)$$

変化検出タスクの損失 入力画像から特徴マップを抽出し、ユークリッド距離より相違度マップを算出する。特徴マップ x はグリッド分割し、 $\|x\| = 1$ となるように正規化したものである。次に、相違度マップ d と教師ラベル T を各グリッドで比較し、変化点検出の損失である L_{Diff} を算出する。算出式を式 (2) で表す。この時、 g は特徴マップをグリッド分割した際の総数である。

$$L_{Diff} = \sum_{i=0}^g \|d_i - T_i\|_2 \quad (2)$$

損失の結合 L_{Seg} と L_{Diff} を足し合わせ 1 つの損失を算出し、学習に用いる。学習に用いる損失 L の算出式を式 (3) で表す。ここで、各損失の重みをそれぞれ w_{Seg}, w_{Diff} とする。重みはあらかじめ定義する。

$$L = w_{Seg} \cdot L_{Seg} + w_{Diff} \cdot L_{Diff} \quad (3)$$

4. 評価実験

本実験では、従来の U-Net と提案手法を用いて、変化点検出の評価実験を行う。

4.1. 実験概要

提案手法では、セグメンテーションと変化点検出の重みを 0.6 と 0.4, 0.7 と 0.3, 0.8 と 0.2 とする。変化点検出の評価実験を行うにあたり、日本の津波被害地域を撮影した Panoramic Change Detection データセット [2] を使用する。データセットは、パノラマ画像ペアで構成される。学習データとして 120 枚、評価データとして 10 枚使用する。また、Data Augmentation を行い、学習データの多様化を図る。評価には F1 スコアを用いる。F1 スコアは、精度と再現率の調和平均であり、式 (4) で求めることできる。

$$F1 = \frac{2 \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

Recall は実際に正であるデータのうち、正であると予測された割合を表し、Precision は正と予測したデータのうち、実際に正である割合を表す。

4.2. 実験結果

図 3 に従来の U-Net との比較を示す。検出結果は、青いほど相違度が低く、赤いほど相違度が高いことを表す。提案手法では、変化した建物の大部分を検出できている。その他、車などの変化した領域が小さい物体に対しても検出精度が向上していることが確認できる。

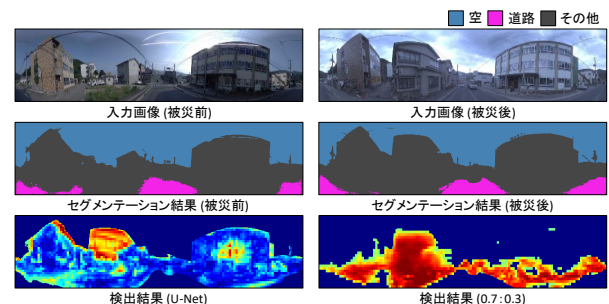


図 3: U-Net と提案手法の比較

表 1 に F1 スコアを用いて評価した結果を示す。U-Net と比較し、タスクごとに分割することで精度が約 0.17 ポイント向上した。また、CNN-feat と比較し精度が約 0.03 ポイント向上していることが確認できる。

表 1: F1 スコアの評価結果

実験条件	U-Net	CNN-feat [2]	提案手法
F1 スコア	0.585	0.723	0.754

5. おわりに

U-Net の Decoder 部分をタスクごとに分割し、変化点検出とセグメンテーションを同時に行う手法を提案した。今後は、学習方法とネットワーク構造の修正を行い、建物の変化をより高精度に検出し、不確かさを用いて、重み付け比を動的に変更する手法について検討する。

参考文献

[1] O. Ronneberger, et al., "U-Net: Convolutional Networks for Biomedical Image Segmentation", MICCAI, 2015.
 [2] K. Sakurada, et al., "Change Detection from a Street Image Pair using CNN Features and Superpixel Segmentation", BMVC, 2015.