

セマンティックセグメンテーションモデル 性能向上レポート

本レポートは、NYUv2データセットを用いたセマンティックセグメンテーションタスクにおいて、mIoU (mean Intersection over Union) 0.50062の精度を達成したモデルに実装された主要な機能と、その改善過程をまとめたものである。

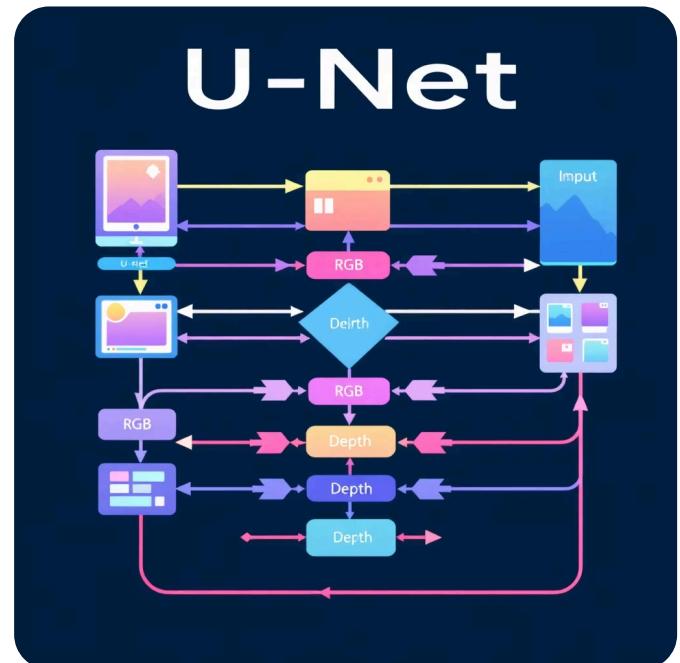
初期のベースライン精度から向上を達成するために、段階的なアプローチでモデル、データ処理、学習プロセスが最適化された。

ベースラインモデルの課題と精度向上に貢献した主要機能

ベースラインモデルの課題

当初のベースラインモデル（シンプルなU-Net構造）は、RGB画像と深度画像を結合した4チャネル入力を利用し、基本的な畳み込み層とReLU、Batch Normalizationで構成されていた。

このモデルは、データ拡張が限定的であり、mIoUは30%台にとどまっていた。主な課題は、モデルの表現力不足、データ処理の不十分さ、および学習プロセスの最適化不足であった。



1

強力なデータ拡張の導入と調整

モデルの汎化性能を高め、過学習を抑制するために、多様なデータ拡張手法が導入された。

- 幾何学的変換：ランダム拡大縮小、ランダム回転、ランダム水平反転
- 画質・色調変換：色調調整、ぼかし、シャープネス調整
- 深度画像およびセグメンテーションラベルの補間モードを
InterpolationMode.BILINEAR から
InterpolationMode.NEAREST へと変更

2

事前学習済みバックボーンの導入

モデルの表現力と特徴抽出能力を飛躍的に向上させるため、

U-Netのエンコーダ部分が ImageNetデータセットで事前学習されたResNet50モデルに置き換えられた。

- `torchvision.models.resnet50`を使用し、事前学習済み重みをロード
- 4チャネル入力に対応するようにResNetの最初の畳み込み層を再構築
- デコーダ部分はResNetの各層からの特徴マップを受け取るように再設計

3

Squeeze-and-Excitationブロックの導入

モデルのチャネル方向のアンシジョン能力を高めるため、各DoubleConvブロックの後にSqueeze-and-Excitation (SE) ブロックが追加された。

- SEブロックは各特徴チャネルの重要度を動的に学習
- 重みに基づいて特徴マップを再調整
- モデルがより情報量の多いチャネルに焦点を当て、特徴表現を最適化

学習プロセスの最適化と堅牢な機能実装



学習プロセスの最適化

学習プロセスの安定化と最適な収束を促すために、以下の最適化手法が導入された。

- Adam オプティマイザ：広範なタスクで実績のある適応的な学習率調整機能
- 学習率スケジューラ (ReduceLROnPlateau)：検証損失が改善しない場合に学習率を自動調整

堅牢な学習再開およびデバッグ機能

一部不安定なクラウド環境での学習を効率的に進めるために、以下の機能が不可欠であった。

- チェックポイントの自動保存：最良のmIoUモデルと定期的な保存
- 最新チェックポイントからの自動学習再開：中断時も効率的に継続
- DataLoaderのデバッグ機能：早期問題特定
- num_workers の調整：データロードの並列化による学習速度向上

① データ拡張の重要な調整ポイント

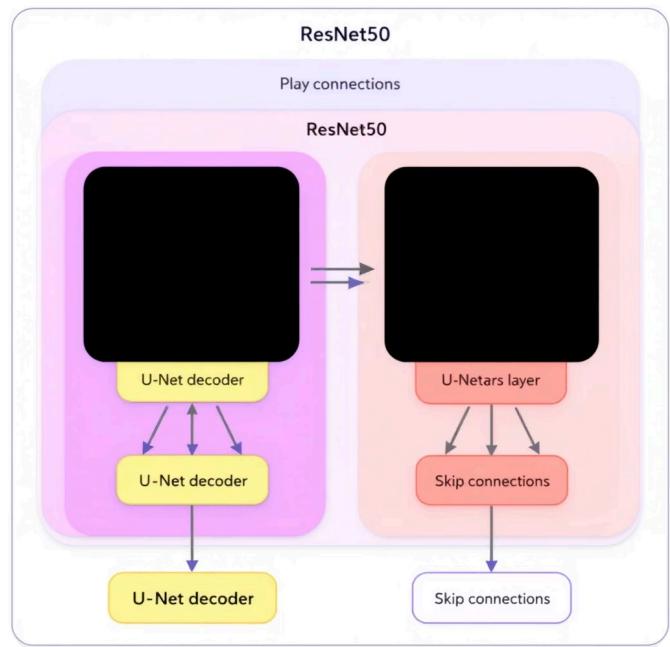
幾何学的変換は、RGB画像、深度画像、セグメンテーションラベルに厳密に同じパラメータで同期して適用された。また、初期は広範囲なスケール変化や回転を許容していたが、精度低下が見られたため、より現実的な範囲 (scale=(0.9, 1.0)、回転-3度から3度) に調整された。色調調整も非常に控えめな強度 (brightness=0.05, contrast=0.05, saturation=0.05, hue=0.02) に設定された。

モデル改良の詳細と結論

ResNet50バックボーンの実装詳細

ResNet50が標準で3チャネル入力であるのに対し、本モデルはRGB（3チャネル）と深度（1チャネル）を結合した4チャネル入力を利用するため、ResNetの最初の畳み込み層（resnet.conv1）が4チャネル入力に対応するように再構築された。既存の3チャネル分の重みはコピーされ、新しく追加された1チャネル分の重みはゼロ初期化された。

デコーダ部分は、ResNetの各層（conv1_bn_relu、layer1、layer2、layer3）からの特徴マップを適切なチャネル数と解像度で受け取り、スキップ接続として結合するように再設計された。デコーダのアップサンプリング層（ConvTranspose2d）とスキップ接続を結合する際には、
`torchvision.transforms.functional.center_crop` を使用して、空間的サイズが正確に一致するように調整された。



ベースラインモデル

シンプルなU-Net構造、限定的なデータ拡張、mIoU 30%台

1

データ拡張の最適化

幾何学的変換と補間モードの調整、画質・色調変換の微調整

2

アーキテクチャの強化

ResNet50バックボーンの導入、SEブロックの追加

3

学習プロセスの最適化

Adamオプティマイザ、学習率スケジューラ、堅牢な学習再開機能

4

最終結果

mIoU 0.50062の達成

5

本プロジェクトでは、シンプルなU-Netモデルから始まり、データ拡張の徹底的な調整、事前学習済みResNet50バックボーンの導入、SEBlockによるアテンション機構の追加、AdamWオプティマイザとReduceLROnPlateauスケジューラの最適化、そして堅牢な学習再開機能の実装を通じて、NYUv2セマンティックセグメンテーションタスクにおいてmIoU 0.50062という精度を達成した。

Made with GAMMA

この成果は、モデルのアーキテクチャ改良と、データ処理および学習プロセスの包括的な最適化が、深層学