深層学習ベースのステレオマッチングに有効な学習データに関する調査

松原太地 ER20084

指導教授:藤吉 弘亘

1.はじめに

深層学習ベースのステレオマッチング手法に, Cascaded REcurrent Stereo matching network (CREStereo) [1] ガ ある. CREStereo は独自に作成した大規模な CG データ セット (CREStereo データセット) を用いて学習すること で、従来手法と比較して高い精度とロバスト性を確保して いる. CREStereo データセットは、CG 環境に様々なオブ ジェクトをランダムに配置して,多様性の高いデータで構 成されている.本研究では、CREStereo データセットを作 成したシーンに焦点を当て、どのようなシーンが精度向上 に寄与するかを調査する.

2. CREStereo データセット

CREStereo はモデルとデータセットの両者に対して精 度向上の観点からアプローチを提案している. モデルは低 解像度から高解像の特徴マップを推論時に階層的に利用す る.低解像度画像の特徴で全体構造を理解し、高解像度画 像の特徴で細部の特徴の抽出を実現している. データセッ トは 200,000 枚の学習データから構成されている. 図1に CREStereo データセットの例を示す. 従来のデータセット と比較して、CREStereo データセットは生物、木、穴があ るものなどの多種多様な物体,照明条件,反射などの実環 境の要素をランダムに含むシ -ンを生成している.



図 1: CREStereo データセットの例

3. データの傾向調査

従来研究では、データセットの改善による精度・ロバスト 性の向上を主張するにも関わらず, データセットの工夫点に ついて十分な議論が行われていない. そこで, CREStereo データセットを要素ごとに分割し,各要素がモデルの学習に 与える影響を調査する. CREStereo データセットは, Hole, Reflective, Shapenet, Treeの4つの要素から構成されて いる.本研究では、データの傾向の調査の手法として、す べての要素で学習,要素を1つずつ除外し学習した結果 を比較して構成要素の学習に与える影響を調査する.評価 データには、ステレオマッチングのベンチマークで広く利 用される屋内と屋外のシーンを多く含む実環境データであ る Middlebury Stereo データセットを利用する.

4.評価実験

本実験では,データセットが学習に与える影響を検証を するために、データセットを分割し、分割後のデータで学 習・推論を行う. 定量的な評価指標には Average Error と Bad を使用する. Average Error は予測した視差と正解画 像間における差の平均値を評価する. Bad はピクセルごと に予測視差と正解画像の値の差を求め、差の絶対値が N 以 上あるピクセルの割合を評価する.

4.1.実験条件

学習用データセットには CREStereo データセット,評 価用データセットには Middlebury Stereo データセットを 用いる. データ分割による傾向調査のために学習データの 4つの構成要素をそれぞれ除外して学習を行う. 最適化手法 は Adam, バッチサイズは 4, イテレーション数は 300,000 回とする.評価指標の Bad の N は 2 とする.

4.2.実験結果

表1に学習データを要素ごとに分割した際の定量的評価 結果を示す.表1より、Reflectiveの要素を含むデータを 除外した場合と全データを学習した場合を比較すると、精 度が約 3.5pt 向上していることが確認できる. Reflective は多様なシーンを表現するためランダムに反射条件を決定 し、図2のような実環境では再現性のない過剰な反射を含

む画像がある.このような再現性のない画像を除外するこ とが精度向上に有効であることが確認できた.加えて再現 性のない画像は特徴マップ上で Reflective の重心付近に多 く分布しており、そこを除外することで精度が向上するこ とを確認できた.また, Shapenet と Hole を除外した際に 精度が約 8.0pt 低下した. これより, 2 つの要素が実環境 データに適しており、CREStereo データセットの精度向 上に大きく影響していると考えられる.以上より,実環境 データにおいて Hole と Shapenet が学習に有効であるこ とを確認した一方で, Reflective の要素は学習に有効でな いことが分かった.

表 1: 定量的評価

学習データ	Average Error \downarrow	Bad 2.0 \downarrow
全データ	30.373	58.883
Reflective 除外	26.791	57.013
Hole 除外	38.355	64.895
Shapenet 除外	38.211	64.123
Tree 除外	30.942	52.373



図 2: Reflective データの例

4.3. 定性的評価

図3に定性的評価結果を示す.図3(b)は正解画像,図 3(c) と図 3(d) はそれぞれ CREStereo データセットを全 て学習した結果, Reflective を除外したデータで学習した 結果である. 図 3(c) および図 3(d) より, Reflective を除 外したデータで推論を行った際に物体の輪郭や背景領域の ぼやけの低減が確認できた. これは, Reflective を除外す ることで、背景領域を含め再現性のない反射条件の画像を 除外し、物体の輪郭や背景領域に対してより有効なデータ で学習することができたからである.





(c) 全データ (d) Reflective 除外 図 3: 定性的評価

5.おわりに

本研究では、CREStereo データセットの傾向調査を行 い、深層学習ベースのステレオマッチングの精度向上につ ながるシーンを調査した.実験結果から、学習データから Reflective を除くことで精度が向上することを確認し、各 データセットを定性的に確認した. 今後はより詳細なデー タの分析を行い、汎化能力と精度を両立するデータの選定 を行う予定である.

参考文献

[1] Jiankun Li, et al., "Practical Stereo Matching via Cascaded Recurrent Network with Adaptive Correlation", In CVPR, 2022