5N-02

ドメイン適応を用いた動作認識のための合成動画像の活用の検討

礒井葉那 †　竹房あつ子 ‡　中田秀基 §　小口正人 †

† お茶の水女子大学　‡ 国立情報学研究所　§ 産業技術総合研究所

1 はじめに

ディープニューラルネットワークの進歩に伴う学習データ不足の問題について様々な議論が行われており, その解決策の 1 つに合成データを利用した学習がある. 合成データには生成が比較的容易であるという利点があるが, 合成データを用いて学習したモデルには, 実データ解析時にデータの分布の違いから解析精度が低下するドメインシフトが起こるという課題がある.

ドメイン適応とは, ドメインシフトに対応するための手法であり, 合成データで学習された分類器を実データに用いる場合にドメイン適応が必要とされることが知られている. ドメイン適応の代表的な手法には, 解析したいデータであるターゲットデータと正解ラベルなどの多くの情報を持つソースデータとを同時にネットワークに入力してデータ間に共通する特徴を学習させる DANN[1] などがある. 文献 [2] では合成データで動画像ドメイン適応を行っている. Kinetics-Gameplay というゲームプレイ動画から作成したデータをソースデータに利用して, ターゲットデータ (Kinetics の 30 のサブクラス) の分類に 17.22% から 27.50% の精度向上を達成した. しかしながら, この精度はラベルを使用してターゲットデータで学習した場合の 64.49% には遠く及ばない.

本研究では, 合成動画像データを活用した高精度な実動画像データ解析の実現を目的とし, 写実的な合成動画像データを作成して学習し, その解析精度を調査した. 実データを用いた解析実験の結果, 合成データのみでの学習でもドメイン適応を用いた学習でも現時点では十分な解析精度が得られず, 改善の余地があることがわかった.

図1　Ochahouse-Syn

図2　Ochahouse-Real

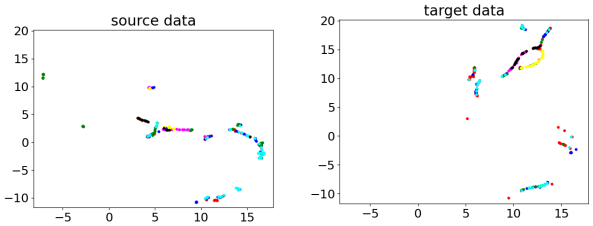

図3　3D ResNet-18 based DANN で抽出した特徴を UMAP で可視化した結果

2 合成動画像データセット Ochahouse Dataset の作成

我々は, 合成動画像におけるドメイン適応のための Ochahouse Dataset を作成した. これは部屋の中を 1 人の人が行動する様子を 1 台の固定されたカメラで収録した合成動画像 Ochahouse-Syn と, 実動画像 Ochahouse-Real により構成される. Ochahouse-Syn の作成には Unity® を使用した. Ochahouse-Real は, 大学の実験住宅 OchaHouse 内で筆者が動作を行い収録した. Ochahouse Dataset では, walking, sitting down, sitting, standing up, lying down, lying, getting up の 7 種類の動作クラスを作成した. 各動作クラスのデータ数は表 1の通りであり, 各動画像は約 3 秒から 7 秒程度の長さとなっている. 作成した動画像データの 1 フレームを図 1, 図 2に示す.

3 実験

作成した Ochahouse Dataset について, ドメイン適応を含む様々な手法で学習したモデルでの Ochahouse-

---

A Examination of Utilization of Synthetic Video Data for Action Recognition using Domain Adaptation

†Hana Isoi
‡Atsuko Takehusa
§Hidemoto Nakada
†Masato Oguchi
†Ochanomizu University
‡National Institute of Informatics
§National Institute of Advanced Industrial Science and Technology (AIST)

表1　Ochahouse Dataset の動作クラスと各データ数

クラス	walking	sitting down	sitting	standing up	lying down	lying	getting up
合成データ Ochahouse-Syn	997	747	1118	780	250	250	250
実データ Ochahouse-Real	96	44	56	51	32	39	32

Real の解析性能を評価する．ドメイン適応を行わない手法では,Ochahouse-Real のみ (target only) または Ochahouse-Syn のみ (source only) を 3D ResNet-18 を用いて学習した．ドメイン適応を行う手法では 3D ResNet based DANN と TemPooling[2], TA^{3}N[2] を用いて学習した．3D ResNet based DANN は, 動画像解析のために, 静止画像のドメイン適応のためのネットワーク DANN[1] の特徴抽出器を 3D ResNet-18 に置き換えたネットワークである．データ拡張としてノイズ・ぼかしの付与と, 明るさ・コントラスト・輝度のランダムな変更を行った．実装にはいずれも PyTorch を用い, source only と target only, 3D ResNet based DANN の損失関数にはクロスエントロピー誤差を, 最適化手法には Adam を, 学習率には 0.001 を採用して産業技術総合研究所の ABCI で 7 クラス動作分類実験を行った．

結果を表 2に, 3D ResNet-based DANN の特徴抽出器で抽出した特徴を UMAP で可視化した様子を図 3に示す．表 2 からわかるように, Ochahouse-Real の正解ラベルを使わないいずれの手法でも 17.14 % から 31.71 % と Ochahouse-Real を高精度に解析することができなかった．3D ResNet-18 based DANN によるドメイン適応により解析精度がわずかに向上したが, 今回の実験では TemPooling[2] および TA^{3}N[2] の効果は確認できなかった．

図 3では, 3D ResNet-18 based DANN で抽出した特徴を UMAP で可視化した結果を示しており, 動作クラスごとに各点が色付けされている．この図から, 学習した特徴抽出器で得られた Ochahouse-Syn と Ochahouse-Real の特徴分布の形状が異なっているが, 動作クラスごとの各点は黒や黄の一部のクラスは集まっていることがわかる．よって, このネットワークはデータ間に共通する特徴を抽出できていないが, クラス固有の特徴をある程度抽出できていることがわかる．この原因として DANN のクラス分類器がドメイン分類器に対し強力すぎることが考えられるため, クラス分類器とドメイン分類器の学習のスケジュールを調整するなど, 各パラメータ調整が必要であることがわかった．

表2　さまざまな学習手法での実データ解析精度

学習手法	精度 (%)
target only	81.14
source only	29.63
3D ResNet-18 based DANN	31.71
TemPooling[2]	17.14
TA^{3}N[2]	28.57

4　まとめと今後の取り組み

本研究では, ラベルなし実動画像データの解析に向けた合成動画像データ活用方法について検討した．まず, 同様なシーンで人間の動作を収録した実動画像データ Ochahouse-Real と合成動画像データ Ochahouse-Syn を作成した．実験の結果, 我々人間の目で見て大きな違いがないこれらのデータ間の違いは今回採用したニューラルネットワークにとっては大きく, 合成データのみでの学習では十分な精度で実データの解析ができないことがわかった．また, ドメイン適応を用いた手法でも現段階では効果が確認できなかった．

今後はドメイン適応を用いた手法でのパラメータチューニングを行い, 効果を確認する．また, オプティカルフローを入力に用いたり, 様々なドメイン適応手法を用いて実験し, 効果的な合成データの活用方法について検討する．

謝辞

この成果の一部は，JSPS 科研費 JP19H04089, JP19K11994 及び，2020 年度国立情報学研究所公募型共同研究 (20S0501) の助成を受けたものです．

参考文献

[1] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P.,Larochelle, H., Laviolette, F., Marchand, M. andLempitsky, V.: Domain-Adversarial Training ofNeural Networks, J. Mach. Learn. Res., Vol. 17,No. 1, p. 2096 2030 (2016).

[2] Chen, M.-H., Kira, Z., AlRegib, G., Yoo, J.,Chen, R. and Zheng, J.: Temporal Attentive Alignment for Large-Scale Video Domain Adaptation, ICCV2019