

弱教師あり学習モデルに対する転移学習を用いた 動画中のイヌの情動認識手法に関する検討

奥村 美生[†] 丸野 由希[†] 久保 孝富[‡] 浦 優輝[‡] 吉村 愛琉[§] 永澤 美保[§]
京都女子大学[†] 奈良先端科学技術大学院大学[‡] 麻布大学[§]

1 はじめに

本研究では、動画中のイヌに対する情動認識システムの開発を目指している。情動認識モデルの学習にはイヌの情動ラベル付き画像データセットの作成が必要となるが、専門家の作業によって大規模なイヌの情動ラベル付きデータセットを得ることは難しい。この問題点を克服するために、比較的規模の大きな犬種ラベル付きデータセットで弱教師あり学習モデルを事前学習させた後、イヌの情動認識に利用することを試みた。本稿では、その基礎的検証内容としてイヌの3種類の情動に対する認識結果を報告する。

2 情動認識システム

本システムには、データ拡張を伴った弱教師あり学習のニューラルネットワークモデルである Weakly Supervised Data Augmentation Network (WS-DAN) [1] を用いる。WS-DAN は、画像分類課題を実行するために、対象物体の各パーツを表現できるように一連の局所的な特徴量を抽出するためのアテンションマップを備えている。このアテンションマップは、データ拡

張の際にも利用されており、データ拡張の有効性改善にも寄与するよう期待されている。

本システムでは、WS-DAN に犬種ラベル付きデータセットで事前学習をさせることで、少量の情動ラベル付きデータセットしか得られない場合であってもイヌの情動認識を実現可能とする。犬種ラベル付きデータセットでの事前学習からイヌの画像特徴を獲得でき、その獲得した画像特徴が情動認識時にも有用であろうと我々は考えている。

3 検証方法

犬種ラベル付きデータセット Tsinghua Dogs dataset [2] を用いて事前学習済みモデルを作成し、そのモデルに転移学習を加えることで、情動認識を行う。情動認識の検証には、著者達が独自に計測・アノテーションした動画データを用いる。一匹のイヌを対象に継続的に動画を撮影したところ、11種類の情動のラベル付けが得られた。そのうち、本稿では、「嬉しい」、「とまどい」、「悲しい」の3種類の情動を用いて検証を行うこととし、関連する場面から5,770枚の画像を抽出した。

4 検証結果

犬種ラベル付きデータセットで事前学習したモデルを使用してイヌの情動状態判別を行ったところ、認識精度は80.33%であった。図1に混同行列 (confusion matrix) を示す。

次に我々は、アテンションマップの確認を行った。犬種ラベル付きデータセットで事前学習しなかったモデルのアテンションマップで

A dog emotion recognition system from videos using a weakly-supervised model and transfer learning

[†] Mimi Okumura and Yuki Maruno, Kyoto Women's University

[‡] Takatomi Kubo and Yuki Ura, Nara Institute of Science and Technology

[§] Haru Yoshimura and Miho Nagasawa, Azabu University

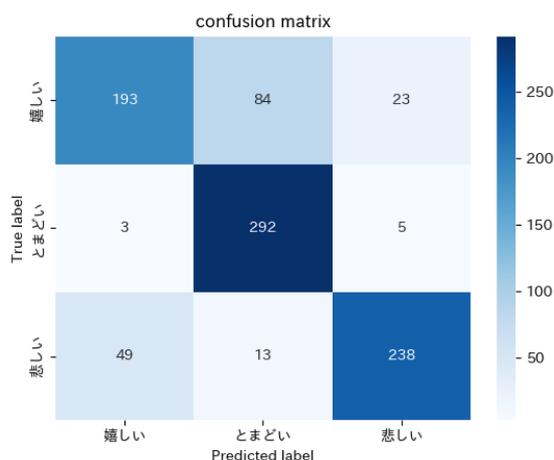


図1: 情動認識結果の混同行列



図2: アテンションマップの1例. 左: 犬種ラベル付きデータセットによる事前学習なし, 右: 事前学習あり.

は、イヌの顔が注視されていないことが分かる(図2左側). 一方で、犬種ラベル付きデータセットで事前学習したモデルのアテンションマップでは、イヌの顔や腕が注視されている(図2右側). また、別の画像の例では、犬種ラベル付きデータセットで事前学習しなかったモデルのアテンションマップではイヌが認識されていないが(図3左側)、事前学習したモデルのアテンションマップでは、イヌが認識されていることがわかる(図3右側). 犬種ラベル付きデータセットで事前学習したモデルを使用し転移学習させることで、場所・背景でなくイヌの表情・姿勢を学習できていると推測され、背景が異なるような新たなデータセットに対しても情動認識が可能であると期待している.



図3: アテンションマップの別の例. 左: 犬種ラベル付きデータセットによる事前学習なし, 右: 事前学習あり.

5 まとめ

本稿では、犬種ラベル付きデータセットで事前学習したモデルに転移学習させることで、イヌの情動認識が可能となるか検討を行った. 3種類のイヌの情動状態を認識させたところ、80.33%と期待が持てる性能が得られた. また、そのアテンションマップでは、背景ではなくイヌの体部位を相対的に注視している傾向がみられたことから、イヌの表情・姿勢を適切に学習できている可能性がある. イヌの情動を認識することは可能であったが、認識率にはまだ改善の余地がある. 今後より多様な個体・情動状態をカバーし、詳細な検討を行う必要があるが、本稿での提案システムはそれに向けて期待が持てる性能を示したと捉えている.

参考文献

- [1] Tao Hu, Honggang Qi, Qingming Huang, and Yan Lu. See better before looking closer: Weakly supervised data augmentation network for fine-grained visual classification. *arXiv preprint arXiv:1901.09891*, 2019.
- [2] Ding-Nan Zou, Song-Hai Zhang, Tai-Jiang Mu, and Min Zhang. A new dataset of dog breed images and a benchmark for fine-grained classification. *Computational Visual Media*, 2020.