モーショントランスファーに基づく演技動作評価法

東京工科大学　バイオ・情報メディア研究科　コンピュータサイエンス専攻

亀田研究室 D2120002 斉玉

1　紹介

本研究は、動物練習[1]の歴史と動物練習課程[1]の現状をまとめた。 動物練習の教授法と評価基準が分析され、論じる。 教育の過程で、学生の自分の内省能力を向上させることは、教育目的をよりよく達成することができると信じている。この観点から、 coco データセット[2]に基づいてサルの手足のキーポイントの QMonkey データセットを作成する。これには、8 本のビデオに 1428 枚の連続した写真が含まれている。 QMonkey データセットを通じて、一部のサルに適したキーポイント検出モデルがトレーニングされ、人間とサルの間の動きの伝達に使用され、立っていないポーズに適した標準化された方法が提案される。また、モーショントランスファーのアクション模倣性に応じた応用方法を提案し、応用過程における大きな可動範囲や複雑なアクションの問題の解決策を提案する。

2　研究目的

本研究では、動物練習課程の教授法と評価基準を分析し、論じる。教育の過程で、学生自分の内省能力を向上させることは、教育目的をよりよく達成することができると信じています。 この観点から、動物練習とモーショントランスファーを組み合わせた新しい教育プログラムが提案される。

モーショントランスファー技術における、画像類似度指標を通じて、演技専門学習者の動物練習課程に客観的な評価基準を提供することを目標としている。

動物練習課程の信頼できる評価方法を提供し、採点における教師の主観的な影響を排除し、人工知能と演劇教育の組み合わせの経験を提供することを望んでいる。

3　問題提出

動物練習課程の現状をまとめて分析する。 この論文は、ロシア、中国、日本、および米国における動物練習課程の設定とパフォーマンス専攻の指導状況を調査する。一般的な動物課程の練習は：観察、練習。、試験 3 つのフェーズに分けられる。動物の練習課程の現状を調査する中で、教育過程でいくつかの問題を発見した。

動物練習課程には統一された評価基準がなく、教授法も異なるため、教師は教科書の動物運動理論と教育経験に基づいてのみ、課程の内容を学生に教えることができる。 このマスターと見習いの間の口頭での教えと、内面の理解な指導方法は、必然的に教育の偏りにつながる。学生は、学習過程において道具に頼りすぎており、動物練習課程で学びたい能力を無視することもある。

研究の過程で、いくつかの問題のため、人と動物の間の移動は非常に挑戦的である：まず、サルという特徴が明らかで、動物の練習課程で比較的に頻度の高い動物を実験主体として選択した。

4　研究背景

4.1　動物練習

動物練習の理論は、ソ連のパフォーマンスアーティスト、スタニスラフスキーが提唱した俳優のトレーニング方法。現在は演技専攻の基礎課程に多く応用している。通常、大学 1 年、学生たちは 16 週間の動物練習課程を行う。動物練習はスタニスラフスキーのシステムに従い、様々な国に導入された。各国は、それぞれの演劇の特徴に合わせて研究開発を行ってきった。動物練習課程の学習内容は、動物の「行、食べる、寝る、狩猟」などの動作を観察して模倣し、練習中、学生たちは賢いサル、獰猛なトラ、攻撃的な雄鶏など、動物の鮮やかな模倣を通じて、肢体の柔軟性を鍛え、舞台での心身のリラックスを実現する。

本論文は、ロシア、中国、日本、および米国における動物練習課程の設定と演技専攻の授業現状を調査する。

動物練習課程には統一された評価基準がなく、教授法も異なるため、教師は教科書の動物練習理論と教育経験に基づいてのみ、課程の内容を学生に教えることができる。 この教師と学生の間の口頭での教えと、内面の理解な指導方法は、必然的に教育の偏りにつながる。

学生は、学習過程において道具に頼りすぎており、動物練習課程で学びたい能力を無視することもある。

学生は教師の評価に頼りすぎて、自分の状況を無視し。一部の学生は、より良い試みをあきらめるだけでなく、一部の学生が自分の限界を達成し、低得点を達成する原因となる。

4.2　モーショントランスファー

ヒューマンモーショントランスファーとは、ソース モーション ビデオ内のオブジェクトのアクションをターゲット オブジェクトに転送し、ターゲット モーション ビデオを生成すること。効果の高いアクショ

[1] 動物練習課程:動物練習課程は、演技専攻の基礎課程。 動物の形態や動きなどを観察し、模倣することにより、身体運動と自然解放の目的を達成する。

ン転送技術は、人間のジェスチャ認識、ターゲット画像の生成、ジェスチャの正規化の 3 つのステップから切り離せない[n-n]。本研究では,[everybody][3]に基づき,人とサル間の動作伝達に適した革新を行った.

4.3 データセット

Coco データセットは、大規模なオブジェクト検出、セグメンテーション、およびキャプション データセット，25,000 の注釈付きボディ画像がある，これには、鼻、手首、膝など、17 のキー ポイントのラベルが含まれる。coco データセットを使用して openpose モデル[4]をトレーニングし、人間の姿勢を識別する。オブジェクト検出データセットの中で、動物関連の姿勢認識データセットは見たことがない。

4.4 画像類似度測定

LPIPS[5]は、画像の類似性を判断するための知覚メトリックとして深度機能を使用するインデックス。 これは、SSIM や PSNR などの広く使用されている構造評価指標とは異なる。 2 つの画像の類似性を評価するために、より人間らしい知覚方法を使用する。 動物摸倣はパフォーマンス能力を向上させる方法であるため、最終的には観客に見せられる。 現時点では、視覚的な類似性が特に重要であるため、動物摸倣の評価基準として LPIPS を選択する。

5 演劇教育における動物練習課程の開発と革新 ---モーショントランスファーに基づく評価方法

5.1 研究方法

本研究はサルという特徴が明らかで、動物の練習課程で使う頻度の高い動物を実験主体として選んだ。実験プロセスは、ターゲットビデオとしてサルのビデオを使用して生成モデルをトレーニングし、ソースビデオとしてサルのアクションを模倣した人間のビデオを使用して、人間のアクションに基づいてサルのビデオを生成する。図 1 に示すように、キーポイント情報のみを含むマカク用の coco 形式のデータセット QMonkey を作成した。現在、データセットの数が少ないため、効果的に識別できるサルはごくわない。後でデータセットを拡張する。

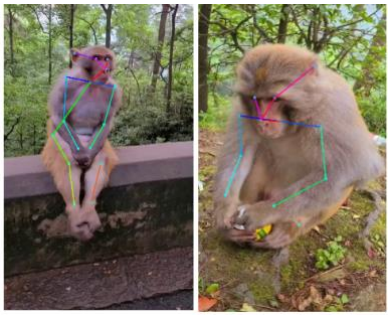

図 1 Qmonkey データセットのイメージ

既存の動物のビデオから、動物とカメラ間の距離が制御不能であることがわかった。 撮影中、彼らは私たちが望む動きをしませんし、設計した範囲内を歩くことはない、小さな動物は、画面全体を埋める可能性があり、大きな動物は、カメラから非常に遠く離れているかもしれない。 そこで、ターゲットサイズの比率の不一致に対して、簡単で一般的な方法を提案する。図 2 に示すように、まず,各フレーム画像におけるサルの姿勢の上下左右 4 つの最も辺点を記録し,サルの平均幅 Mwidth と高さ Mhigh を算出し,人間の幅 Pwidth と高 Phigh を同理計算し,次に下記式により充填すべき割合を算出した.

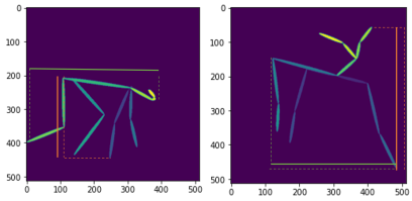

図 2 計算の例

Wscale = (Mwidth/ Pwidth) -1

Hscale = (Mhigh/ Phigh) - 1

Wscale と Hscale は負の値でない場合は,サル画像の塗りつぶし尺度として大きな値を選択する. Wscale と Hscale の両方が負の場合,小さい数値を絶対値としてサル画像の縮小割合として選択する.

図 3 示すように、本研究の方法は、everybody の方法よりも動物と人間の間の正規化に適している。人間がサルを模倣する図とサルの図に基づくモデルでは,本研究の手法と everybody の手法がそれぞれ姿勢図を生成し,我々の手法では姿勢図のサイズが原図に近い.everybody の方法は、サイズが大きすぎると関節の一部が欠落してことがある。

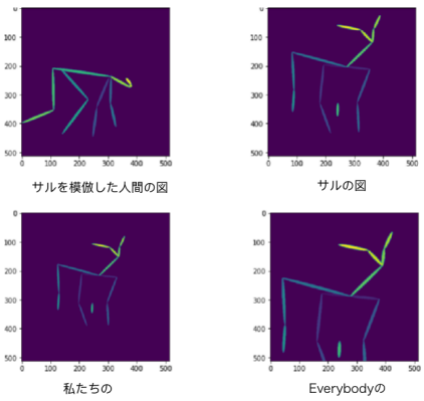

図 3 本研究のメソッドと everybody メソッドの比較図

図 4 は、2 組の比較プロットを使用して評価方法を構築する。 最初のグループは、人間とサルの姿勢図であり、関節の屈曲の程度が類似しているかどうか、各四肢の方向を正確に反映することができる。 しかし、人間とサルの四肢の比率が異なるため、比較の 2 番目のセット、サルの元の画像、および生成された画像も必要。 これにより、2 つの画像セットの類似性データが得られる。

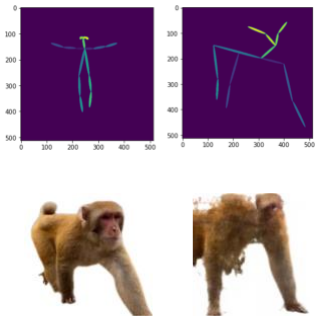

図 4　人とサルの姿勢図とサルの元の画像と生成された図

5.2　研究結果

50 組の比較画像を採点してもらうために、20 人の演技専攻教師を見つけた。比較画像の各グループには、サルの元の画像と、サルの動作を模倣した人間の画像 10 枚、合計 510 枚の画像が含まれている。教師たちはそれぞれ模倣者の肢体動作、歩行規則などの面から採点を行い、評価は高い順に S、A、B、C、D となった。上記および本研究の計算方法に基づいて、これらの実験画像の姿勢図および生成図をそれぞれ数値計算し、平均値を取った。これらの画像の採点比較を分析することで、点数区間が得られた。

実験結果によると、教師が与えた成績が S である模倣画像の数値はすべて 0.3 未満であり、教師が与えた成績が D である模倣画像の数値はすべて 0.4 より大きい。この数値区間に基づいて、私はこの採点基準を得て、複数の演技専門教師のコメントに基づいて、姿勢図と画像を生成する数値を利用して、動物練習の課程で、学生の日常学習と試験の採点に簡潔な参考基準を提供することを望んでいる。

動物摸倣の毎日の演習とテストスコアでは、教師は 2 つの比較データに基づいて、より客観的に学生を採点することができる。 一連の実験的な比較を通じて、学生を採点するために、骨とのコントラストの数値を平均化する評価基準を導き出している。 その後も改善が行われる。

数値	評価
未満 0.3	S
0.3−0.33	A
0.34−0.37	B
0.37−0.4	C
0.4 以上	D

表 1 評価基準

6　這う姿勢でのモーショントランスファー

6.1　問題提出

以上より、モーショントランスファーとは一般的に人のモーショントランスファーを意味し、元のビデオ中の人物の動作をターゲットビデオの人物に遷移させ、ターゲット人物が元の人物の動作で運動するビデオを生成することを知っている。これは非常に興味深い技術であり、俳優が難しい動作を完成できない場合には、モーション トランスファーを使用してアクターがショットを完成できるようにすることを想像できる。もちろん、現在のモーション トランスファー技術はまだこのレベルに達せず、多くの要因が他の分野での応用を制限している。例えば、効果的なモーショントランスファーモデルを訓練するには、単純で無地の背景を使用し、キャラクターはできるだけ同じ位置に固定し、目標人物の動きデータを大量に必要とし、元のキャラクターの動作をできるだけ最大限に再現などが必要である。既存の研究方向の多くは、元人物と目標人物の肢体関係を是正し、より明確な画像の移行を生成することに力を入れており、上記の問題を解決することはできない。

現在のモーショントランスファーは、元の人物と目標の人物の動作が接近してこそ良い生成効果があるため、模倣に基づくモーショントランスファーのようなものだと考えられている。

動物練習では、学生と教師の間の相互模倣も、学生が動物の動作を学ぶのも、最終的には模倣行為である。モーショントランスファーを使用して、模倣が類似している場合、生成される画像は明確になる;模倣が類似していない場合、明確な画像を生成することは難しい。

6.2　研究方法

一般的な動作遷移訓練データには関節のデータ情報が含まれていないか、各関節を区別せず、腕や脚などの骨格だけをマークして訓練モデルにデータの支えを提供しているが、これは複雑な動作の訓練には十分ではないと考えられている。そこで、ボーン間の鋭角の度数を基準にして、ジョイントにデータを充てんします。隣接するボーンの 3 つの座標を使用して角度を計算すると、ジョイントでの塗り潰し円の半径が大きくなる。肩、肘、膝、股間の 8 つの関節をマークし、体幹の長さの 4 分の 1 を頭部の半径として頭部を図 5 に示すように描きました。最後に、姿勢図の首を背景画像の中心に固定し、首を基準に姿勢図全体を描く。

現在進行中の研究では、従来の研究におけるグローバル標準化とモーションリダイレクトを異なるデータ処理方法を提案する。すべてのポーズデータを同じ位置に固定し、トレーニングデータのモーショントランスファー後の画像として関節角度情報を追加する。より良い学習効果を達成するために、比較方法で模倣の類似性の参照を学生に提供し、より直感的に自分の行動と模倣ターゲットの違いを発見する。

図 5　関節角度情報

7　結論

生成ネットワークが実際の画像を生成できるかどうかは、アクションの模倣が同じかどうかだけでなく、ターゲット間の四肢の比率が類似しているかどうかにも依存することがわった。 生成ネットワークは、異なるサイズの四肢を長くまたは短縮して生成できるが、生成された画像の品質にも影響する干渉耐性を有する。 ここでは,目標を同じサイズで,四肢の割合の違いによる影響を極力除去するために,目標スケール正規化手法を用いた. しかし、カメラアングルによる四肢のスケール変化の問題を克服することはできない。 そこで、キーポイントプロットと生成プロットを比較し、比較結果の信頼性を向上させる。

モーショントランスファーの適用シナリオについて議論し、モーショントランスファーは本質的に模倣行為の生成であり、模倣が似ているほど生成効果が高いと考える。したがって、動物練習課程で学生が結果を模倣するためのフィードバックを提供するために使用するのに適している。 今回の研究により、移動状態や這い上がり状態におけるモーショントランスファーの問題が解決されましたが、今後は、動物と人間の間のモーショントランスファーをさらに研究し、この技術をできるだけ早く教育に適用できるようにする。

参考文献

[1] Stanislavski C. An actor prepares[M]. Routledge, 1989.

[2] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." European conference on computer vision. Springer, Cham, 2014.

[3] Chan C, Ginosar S, Zhou T, et al. Everybody dance now[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 5933-5942.

[4] Cao Z, Simon T, Wei S E, et al. Realtime multi-person 2d pose estimation using part affinity fields[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7291-7299.

[5] Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

発表実績

1，Qi, Y.; Zhang, C.; Kameda, H. HISTORICAL SUMMARY AND FUTURE DEVELOPMENT ANALYSIS OF ANIMAL EXERCISE. In Proceedings of the ICERI2021 Proceedings. IATED, 2021, 14th annual International Conference of Education, Research and Innovation, pp. 8529‑8538.

2，Qi, Y.; Zhang, C.; Kameda, H. Animal Exercise: A New Evaluation Method. Journal of Computer Science Research 2022, 4, 24‑30.