# Latent Action Representations Under Domain Shift in Adroit Hand Manipulation Tasks

Assignee Research

June 7, 2026

**Abstract**

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: What is the impact of domain shift on the cross-task transferability of latent action representations learned by CLAM versus token-based methods when evaluated on the Adroit hand manipulation. 14 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Decoupled Doubly Contrastive Learning for Cross Domain Facial Action Unit Detection. Research question: What is the impact of domain shift on the cross-task transferability of latent action representations learned by CLAM versus token-based methods when evaluated on the Adroit hand manipulation benchmark suite?.

## 2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.5/10.

## 3 Results

11 papers retrieved. 14 claims extracted; 0 independently verified. Quality review score: 3.5/10.

## 4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

## 5 Extracted Claims

| Claim | Verified | Confidence |
|---|---|---|
| The optimal setting for $\lambda$, $\gamma$1, $\gamma$2, $\gamma$3, $\gamma$4 was set as 0.1, 1.0, 5.0, 0.1, 1.0 according to the cross-domain AU detection | × | 0.13 |
| We set hyper-parameter $\tau$ = 0.07 and $\alpha$ = 0.1. | × | 0.02 |
| We implemented all the experiments using PyTorch on two RTX 3090 GPUs, each with 24 GB memory. | × | 0.02 |
| We set the respective training batch size N = 128 for the source and target images. | × | 0.05 |
| We set the learning rate as 0.001 and trained D2CA for 50 epochs until convergence. | × | 0.04 |
| The shared AU encoder Eau has the same network structure with the private domain encoders. | × | 0.05 |
| The encoder is composed of eight convolutional layers, containing 0.35 million parameters, and requires 82.59 million fl | × | 0.01 |
| The inference time is 2.8 milliseconds. | × | 0.02 |
| The image discriminator Ds/Dt in Fig. 6 is patchGAN-based [12]. | × | 0.02 |
| The image decoder G in Fig. 7 consists of seven convolutional layers and an auxiliary branch that repeatedly incorporate | × | 0.04 |
| D2CA achieves an F1 score of 35.5 for AU1, 42.7 for AU2, 50.6 for AU4, 34.0 for AU6, 70.3 for AU12, and 46.6 on average | × | 0.03 |
| D2CA achieves an F1 score of 54.6 for AU1, 56.4 for AU2, 46.2 for AU4, 71.5 for AU6, 67.1 for AU12, and 59.2 on average | × | 0.03 |
| D2CA achieves an F1 score of 45.4 for AU1, 36.5 for AU2, 59.4 for AU4, 39.9 for AU6, 67.2 for AU12, and 49.7 on average | × | 0.03 |
| D2CA achieves an F1 score of 19.2 for AU1, 15.7 for AU2, 14.7 for AU4, 76.6 for AU6, 80.0 for AU12, and 76.2 on average | × | 0.03 |

## References


- http://arxiv.org/abs/2503.08977v1
- http://arxiv.org/abs/2505.04999v1
- http://arxiv.org/abs/2502.00396v2