# Exploring the Link Between Muscle Activity and Human Motion Using Computer Vision

Dhruv Sharma, Birudugadda Srivibhav
Department of Chemical and Computer Science And Engineering
**Project Supervisor: Prof. Shanmuganathan Raman**
**Department Of Computer Science And Engineering, IIT Gandhinagar**

## Abstract

Human motion is created by, and constrained by, our muscles. We are utilizing computer vision methods to delve into this intricate relationship, aiming to represent the internal muscle activity that causes motion. We used a dataset, **Muscles in Action (MIA)**, to learn to incorporate muscle activity into human motion representations. Using this dataset, we learn a bidirectional representation that predicts muscle activation from video, and conversely, reconstructs motion from muscle activation. Putting muscles into CV systems will enable richer models of virtual humans, with applications in sports, fitness, and AR/VR. This innovation paves the way for a deeper understanding of human movement dynamics and opens doors to diverse applications across various industries.

## Introduction

Human motion is intricately linked to muscle activity, controlled by signals from the brain to nerves, causing muscle contractions, and ultimately joint movement. We capture internal muscle activity from video inputs, discerning muscles engaged during various actions. Motion is a delicate balance of muscle forces and gravity, manifesting differently in different phases of actions like squats.

Our proposed framework aims to provide a comprehensive understanding of this intricate relationship, offering insights into how muscle activity influences and shapes human motion. By modeling the association between human motion and internal muscle activity, we envision a paradigm shift in how we analyze and interpret human movement, paving the way for advancements in fields ranging from biomechanics to rehabilitation sciences

## Dataset Description

- The dataset presents **12.5 hours of meticulously synchronized video and sEMG signals.** Signals were meticulously collected to encompass the activity of **eight crucial muscles**: biceps, laterals, quadriceps, and hamstrings, each shedding light on the total bioelectric energy at neuromuscular junctions.
- The dataset has 15 exercises, each executed for a standardized duration of 5 minutes by all 10 subjects, a deliberate effort was made to capture variations in execution speed, exertion levels, and body orientations, thus mirroring the nuanced complexity of real-world physical activities.
- Subjects, are equitably distributed between genders, and each contributing 75 minutes of data, the dataset affords a comprehensive panorama. Furthermore, the inherent variations in subjects' weight and muscle composition constructs a meticulously curated, representative sample, primed for thorough analysis and modeling.

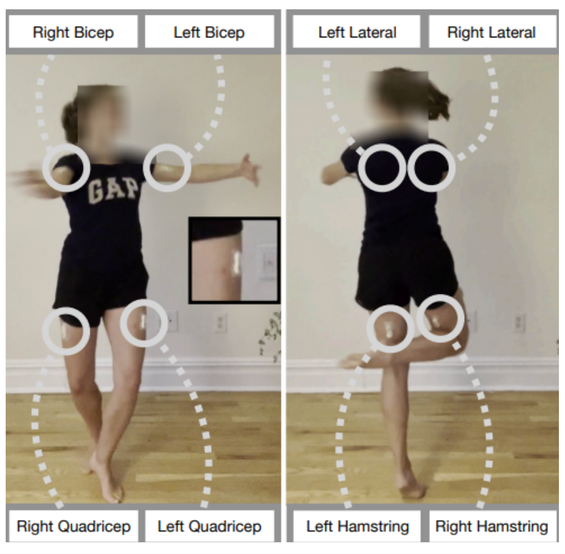



**Figure 1: Sensor Placement.** Illustration of the placement of our 8 sEMG sensors on a subject. We label the 8 measured muscles.

## Data Pre-processing

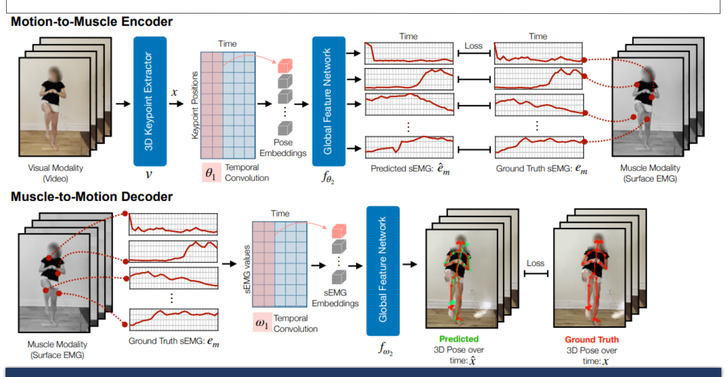
1.Resampling aligns sEMG (10fps) and video (29.97fps) timestamps for modalities synchronization.

2. **VIBE model** extracts **3D/2D keypoints.** 3D normalized, 2D absolute to frame dimensions.

3. Default input: 30 frames; output: 30 sEMG values/muscle (3 seconds). **Train/test split: 20%** randomly allocated to test set; remaining **80%** to training set per exercise per subject.

## Muscle and Motion Mapping

- We aim to establish a **bidirectional mapping** between **visual data** (human poses from video) and **muscle activity** (sEMG signals).
- Human poses are represented as $x \in R^{KD \times T}$,where K is the number of key points, D is the dimensionality, and T is the number of frames. Muscle activity is represented as $m \in R^{MD \times T}$, where M is the number of individual muscles.
- Define two functions:
  **Encoder** (Eθ): Converts human poses (x) to muscle activity (m).
  **Decoder** (Dω): Reconstructs human poses (x) from muscle activity (m).
- Parameters for both models (encoder and decoder) are learned through **supervised learning**, minimizing a mean squared loss between predictions and ground truth in both modalities.
- **Stochastic gradient descent** with the **Adam optimizer** is used for optimization.




## Model Architecture

The described architecture comprises a **common framework** for both the encoder (Motion-to-Muscle) and decoder (Muscle-to-Motion), with minimal adaptations for input and output modalities. Key components include local feature extraction and a global feature network implemented through a Transformer.

**Common Architecture Overview**

Encoder $(E(x) = f_{\theta_2}(g_{\theta_1}(x)))$

Decoder $(D(m) = f_{\omega_2}(g_{\omega_1}(m)))$

**Local Feature Extraction**

For both encoder and decoder-

$(g_\theta(x) = \theta * x)$ For Key points(X)

$(g_\omega(m) = \omega * m)$ For muscle activations(m)

Outputs a sequence of embeddings $(d_1, ..., d_T)$

**Global Feature Network**

Transformer with 4 layers, 8 attention heads, no masking. Input $(d_1, ..., d_T)$ Output $(o_1, ..., o_T)$

**Output Mapping**

Motion-to-Muscle encoder Maps $(o_t)$ to $(m_t)$ sEMG

Muscle-to-Motion decoder Maps $(o_t)$ to $(m_t)$ keypoints

Dimensionality 1) Motion-to-Muscle:- R^M 2) Muscle-to-Motion:- R^KD

## Conditioning and Baselines

We construct conditional encoder and decoder versions, accommodating subject-specific variations. Additionally, we investigate baselines such as Retrieval and Conditional Retrieval to **gauge model effectiveness** and identify potential limitations.

**Conditional Encoder and Decoder:**

Additional encoder and decoder versions adapt to motion, style, sensor placement, and morphology variations with subject-specific conditioning using concatenated tensor *y*.

**Experiment Objective:**

Analyze visual-muscle alignment across 15 exercises using RMSE as the metric for both Motion-to-Muscle and Muscle-to-Motion tasks.

**Baselines:**

Retrieval (Retr.): Nearest neighbor assigns muscle activation or 3D pose of the nearest neighbor to test samples.
Conditional Retrieval (C-Retr.): Similar to Retrieval but conditioned on the subject, aiding comparison with proposed model.

**Baseline Uses:**

Serve as performance benchmarks.
Offer insights and identify potential limitations.
Guide model development and improvement strategies.

## Results

**Quantitative Evaluation**
Both conditional and non-conditional versions of the encoder and decoder are evaluated.
Results are reported per exercise, indicating performance on individual tasks.

**Motion-to-Muscle Encoder**
Both conditional and non-conditional models outperform retrieval baselines for all 15 exercises.
Out-of-distribution experiments show that the learning method generalizes better to unseen exercises compared to retrieval baselines.
Conditional models consistently outperform non-conditional models, confirming a hypothesis presented earlier.

**Muscle-to-Motion Decoder**
Both conditional and non-conditional models outperform retrieval baselines for all exercises.
The conditioned model's performance is less distinguishable from the non-conditioned model, possibly due to inherent conditioning in muscle activity.

**Temporal Analysis**
Seven transformer models are separately trained to predict muscle activation from motion for different input/output lengths.
Both conditional baseline and conditional model improved performance with longer temporal lengths, but the conditional baseline drops in performance beyond 25 frames.

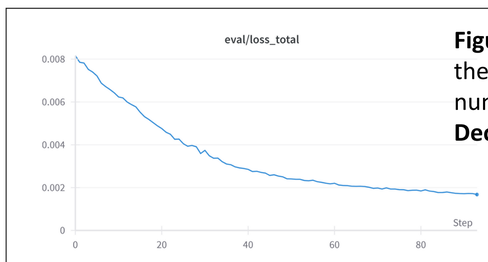



**Figure 2.** Image shows the training loss vs number of epochs for **Decoder Model.**

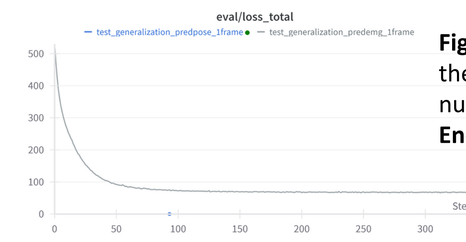



**Figure 3.** Image shows the training loss vs number of epochs for **Encoder Model.**

## Applications

The bidirectional model customizes motions by adjusting muscle activity, catering to individual needs. It optimizes workout effectiveness and minimizes strain, enhancing tailored fitness regimens.

Targeted adjustments fine-tune muscle engagement, optimizing workout efficiency. Selective scaling modifies muscle activation , aligning with desired fitness goals.

The decoder transforms edited muscle activation into personalized workout recommendations. Users receive actionable guidance, maximizing effectiveness and minimizing injury risk.

## References

[1] Mia Chiquier, Carl Vondrick(2023). Muscles in Action. *Columbia University,* https://musclesinaction.cs.columbia.edu/, 2023

[2] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[3] Relja Arandjelovic and Andrew Zisserman. Look, listen and
learn. In Proceedings of the IEEE International Conference on Computer Vision, pages 609–617, 2017.

## Acknowledgements

We extend our gratitude to:
▪Prof. Shanmugnathan Raman for his continuous guidance, support and encouragement throughout the project.
▪Prof. Carl Vondrick, Columbia University, for his guidance in model architecture.
▪M.Tech Student Karan Kumar for his support in lab and assistance with data collections and research suggestions.
▪PhD Scholar MIA Chirquier for her support in understanding model implementation.
▪The staff of the Computer Vision and Human Centered Robotics labs.