Robotics Reading Group
@ Instituto Superior Técnico

Session #8
14-02-2020

Miguel Vasco

"What I cannot create, I do not understand."

Miguel Vasco

“What I cannot create, I do not understand.”

~~Miguel Vasco~~

Richard Feynman

Humans have **rich** generative capabilities

Internal Representations
of External Reality

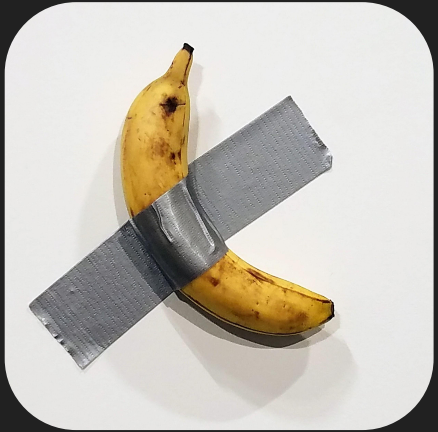
Blindman's buff - a children's game

Humans have **rich** generative capabilities

Internal Representations
of External Reality

Blindman’s buff - a children’s game

External Representations
of Internal Reality



“Banana” - Maurizio
Cattelan (2019)

Humans are able to **interpret** representations

“Accepted with minor revisions”

Watching a scary movie

Warm day at the beach

Last day of school

Humans are able to **interpret** representations

Last day of school

Watching a scary movie

Warm day at the beach

“Accepted with minor revisions”

Humans are able to **generate rich interpretable** motions

Sad

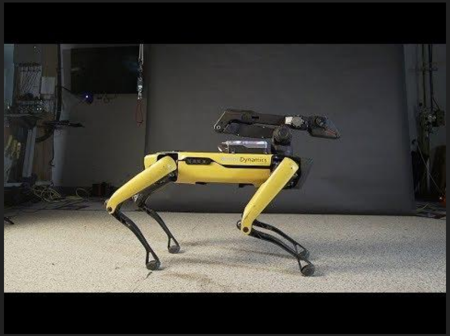
Happy

Robots are able to **reproduce** rich motions

Robots are able to **reproduce** rich motions

Which:

- Are often **Preprogrammed** or hand-animated;
- Require **hours of work** done by animators;
- Can be **repetitive** and **predictable**



Robots are able to **reproduce** rich motions

Makes it difficult to maintain **sustained human-robot interaction:**

Preprogrammed animations

Novelty effect

Conveys lack of intelligence

Stunts long-term interaction

Interaction between two artificial agents with hand-animated motor behavior.

MoveAE: Modifying Affective Robot Movements Using Classifying Variational Autoencoders

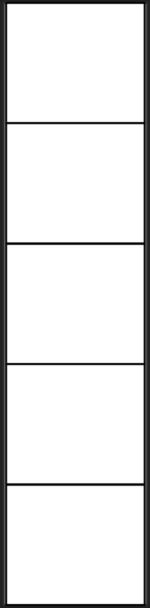
Michael **Suguitan**, Randy Gomez, Guy Hoffman

(Will be) Presented at **HRI'20**, March 23-26, Cambridge, United Kingdom

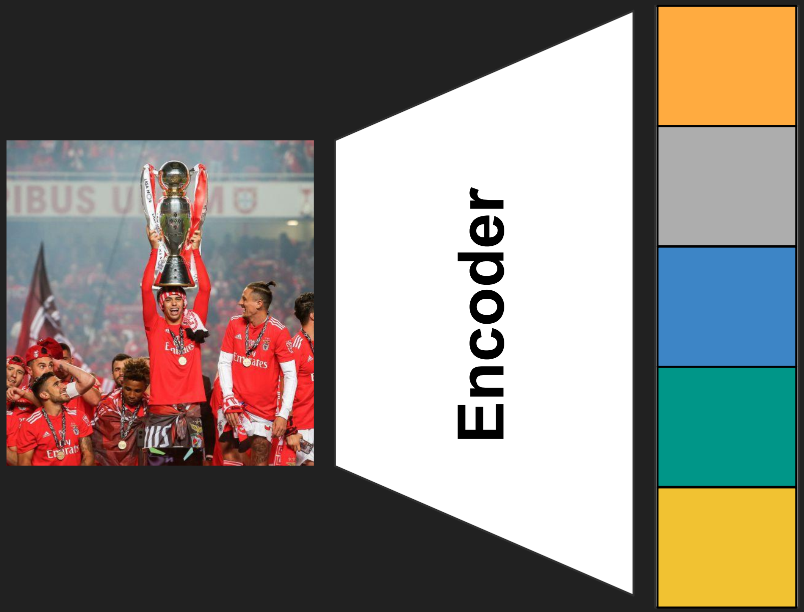
Major Contributions

- A classifying variational autoencoder **architecture** to **reconstruct and generate** expressive **robot movements**.
- A method to **map** the latent space **representations** into the **circumplex emotional model** of valence and arousal.
- An **algorithm to modify** the valence and arousal of movements.

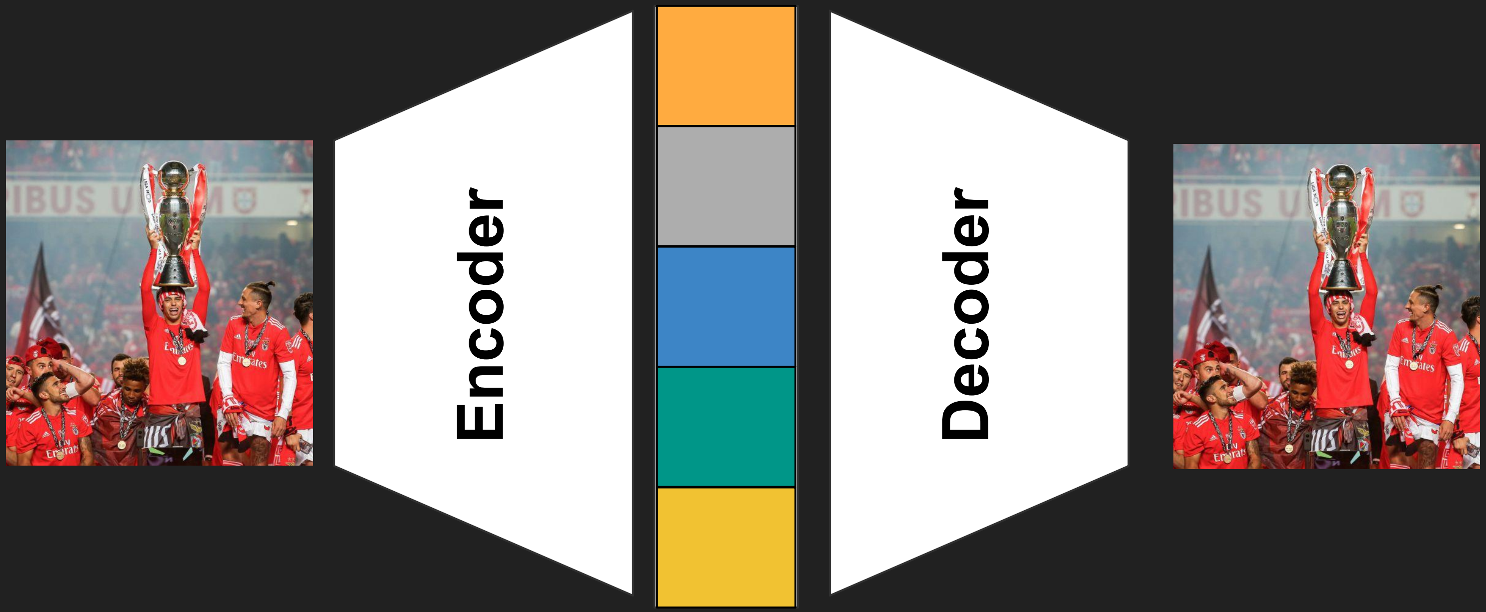
Variational Autoencoder 101



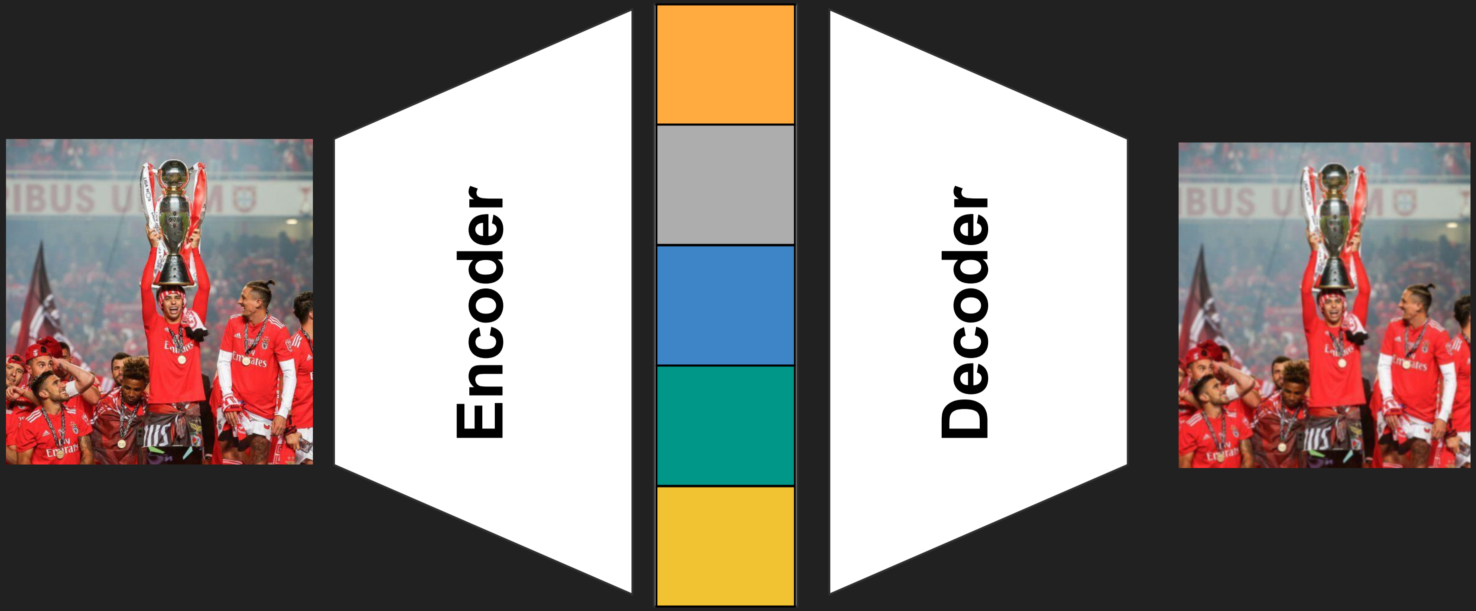
Variational Autoencoder 101



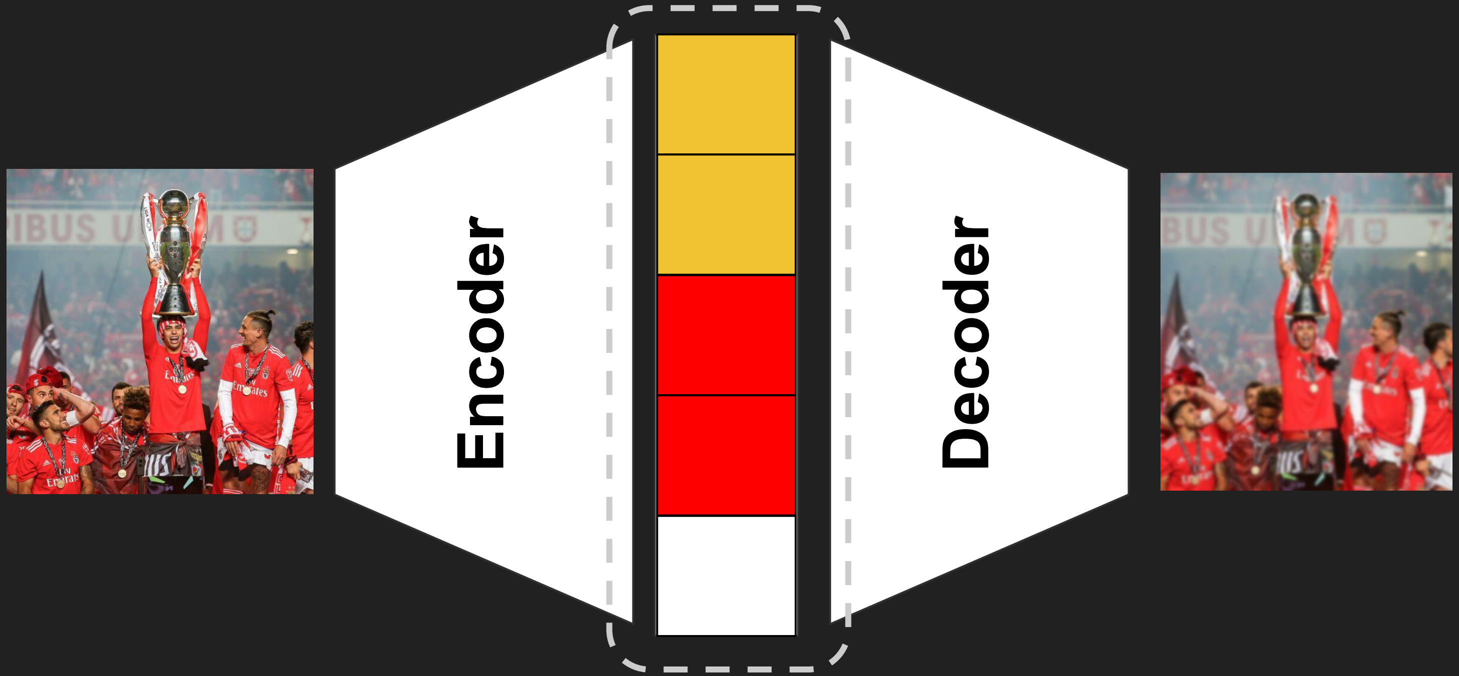
Variational Autoencoder 101



Variational Autoencoder 101

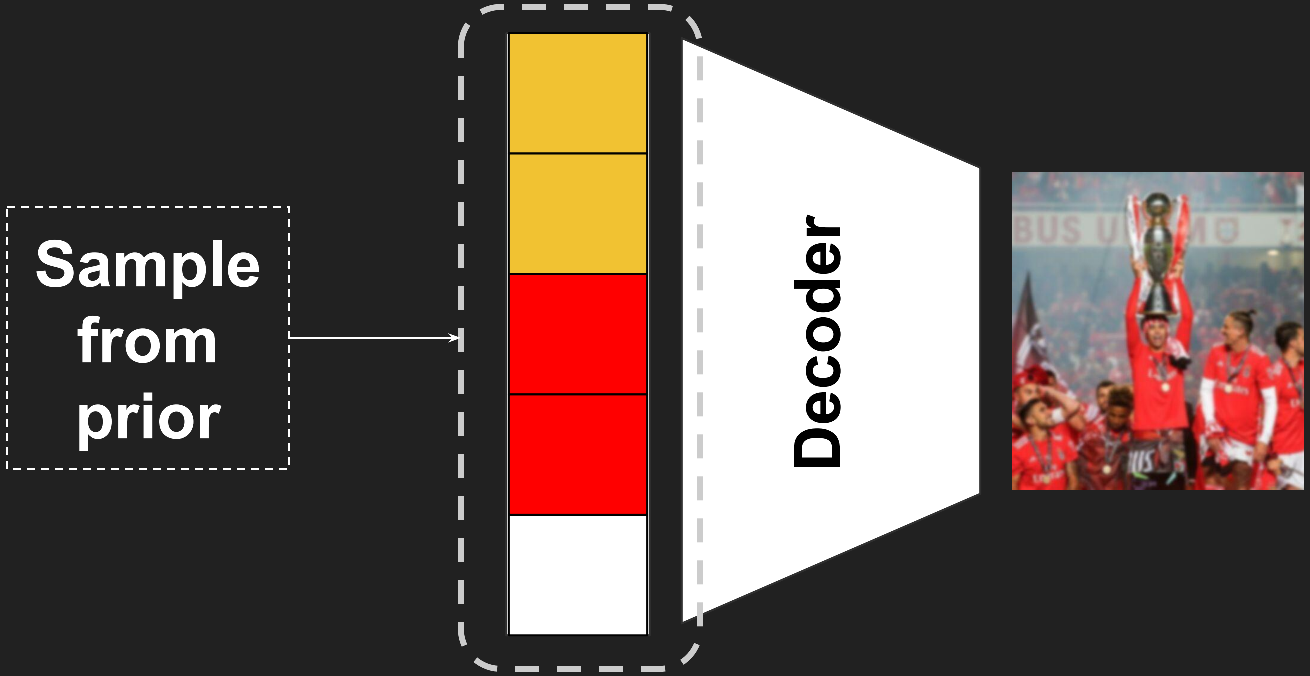


Variational Autoencoder 101

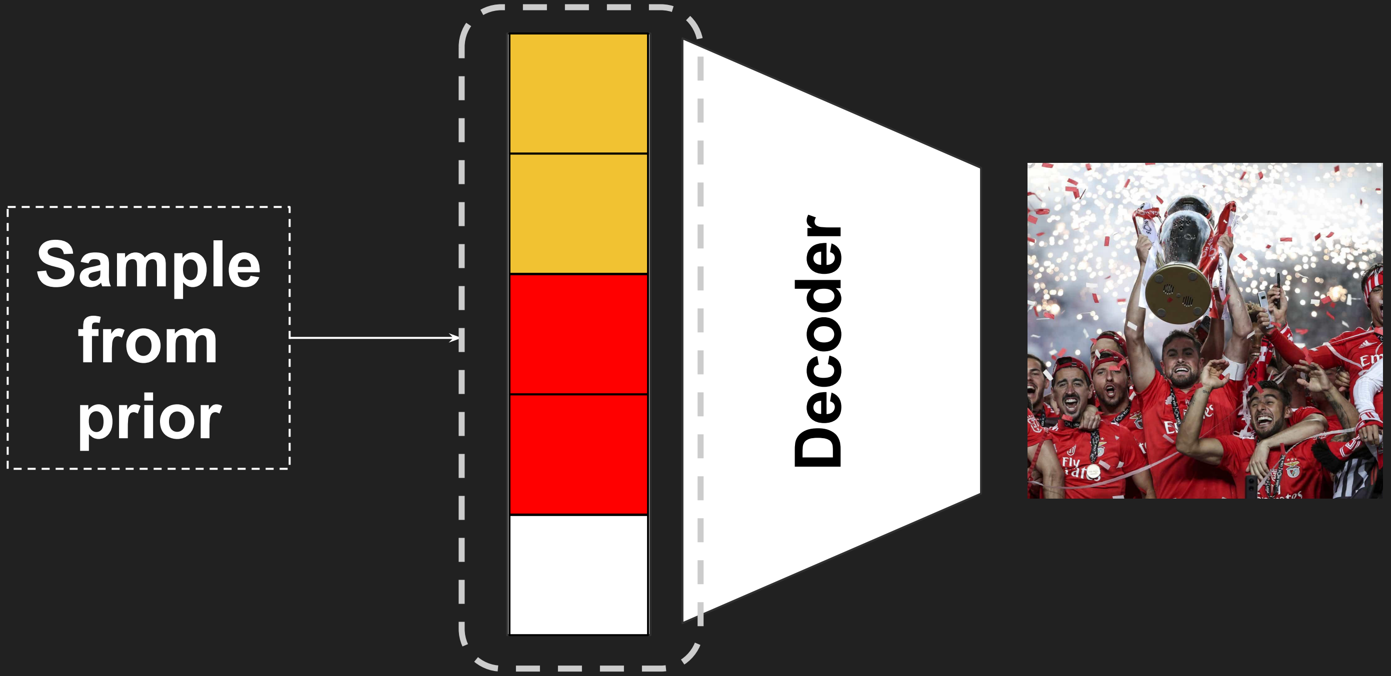


Regularization = Close to prior distribution

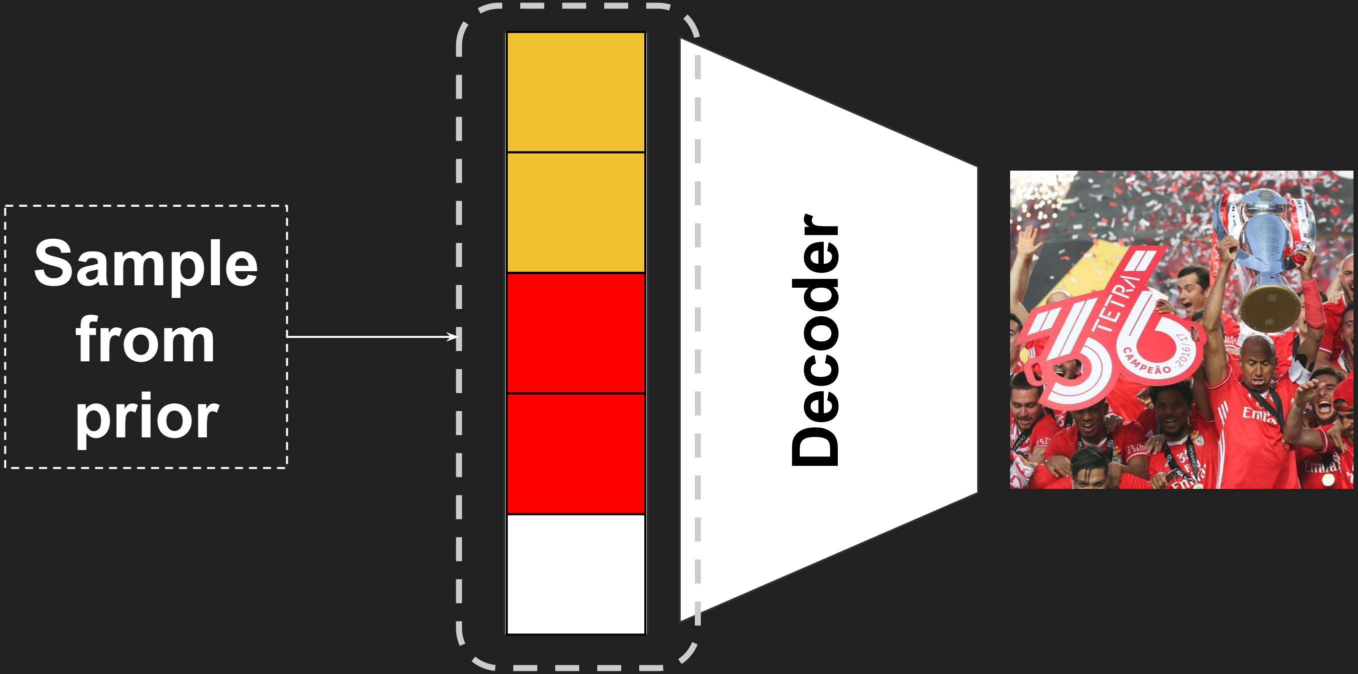
Variational Autoencoder 101



Variational Autoencoder 101

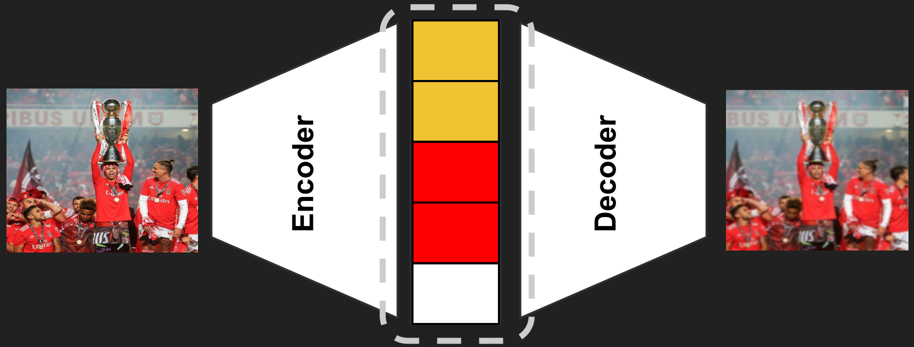


Variational Autoencoder 101

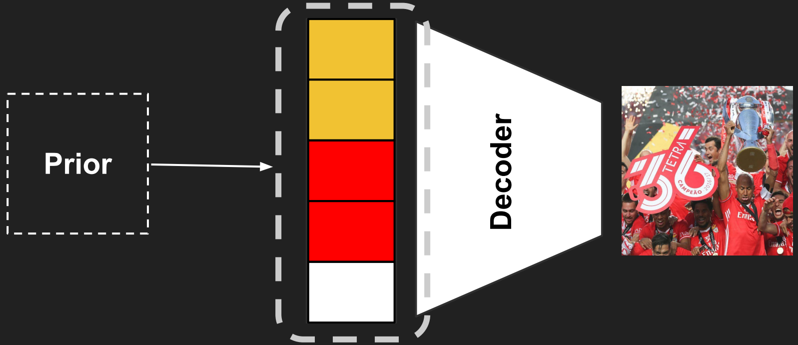


Variational Autoencoder 101

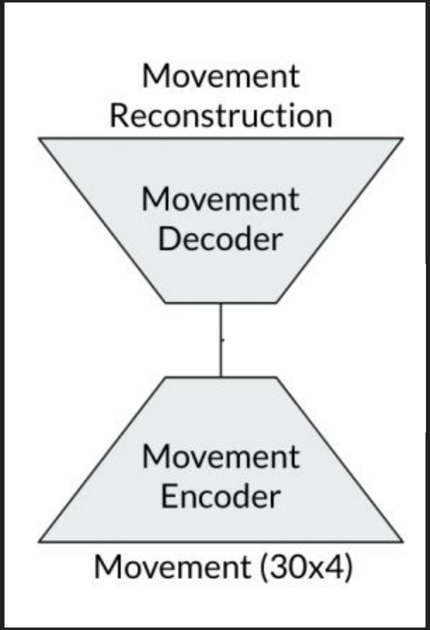
- Reconstruction



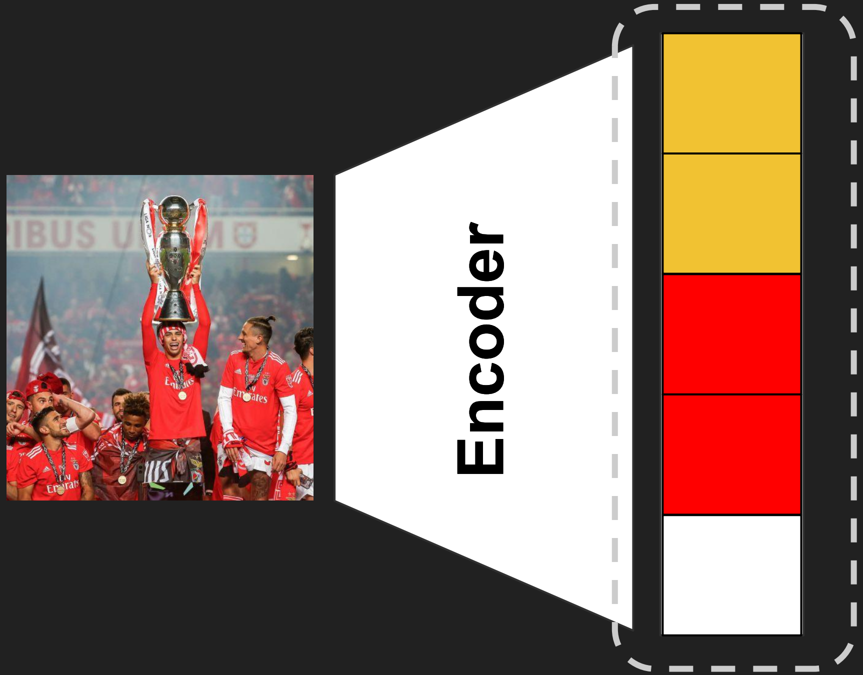
- Generation



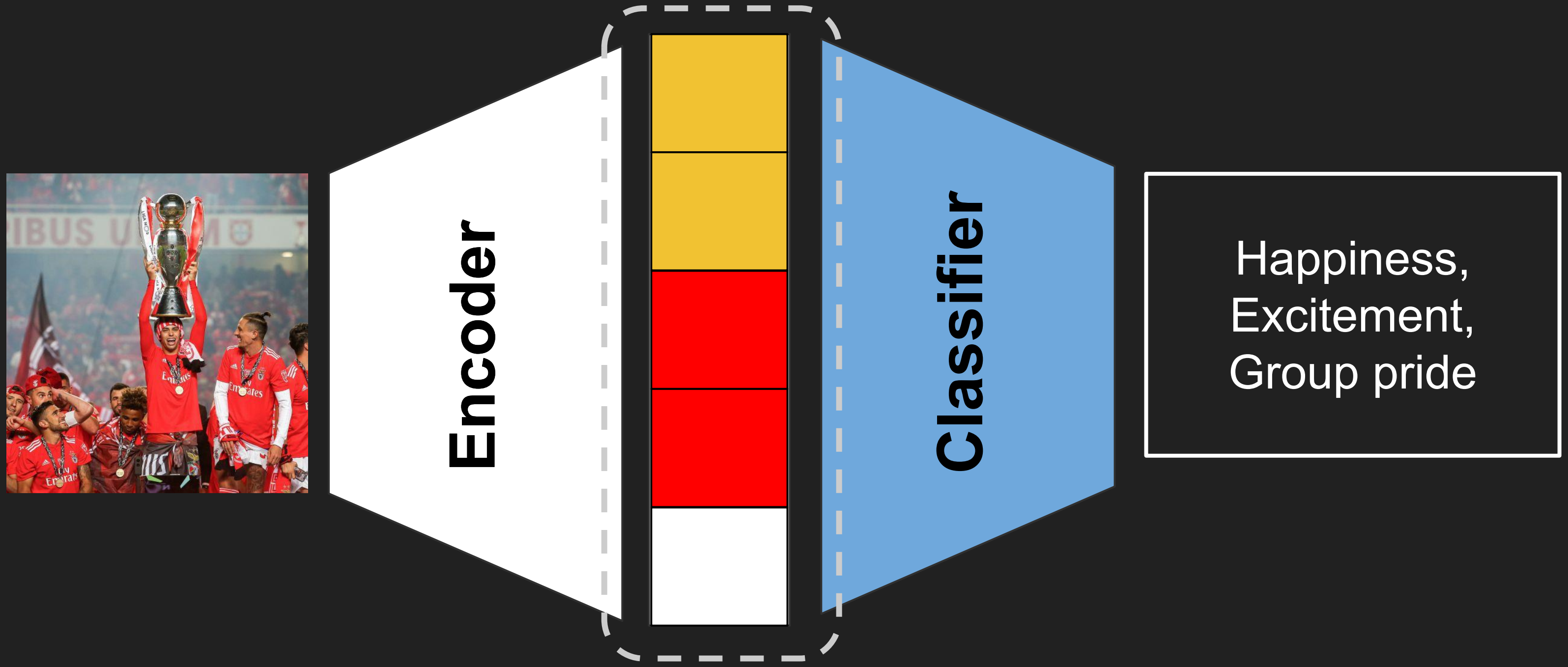
Variational Autoencoder



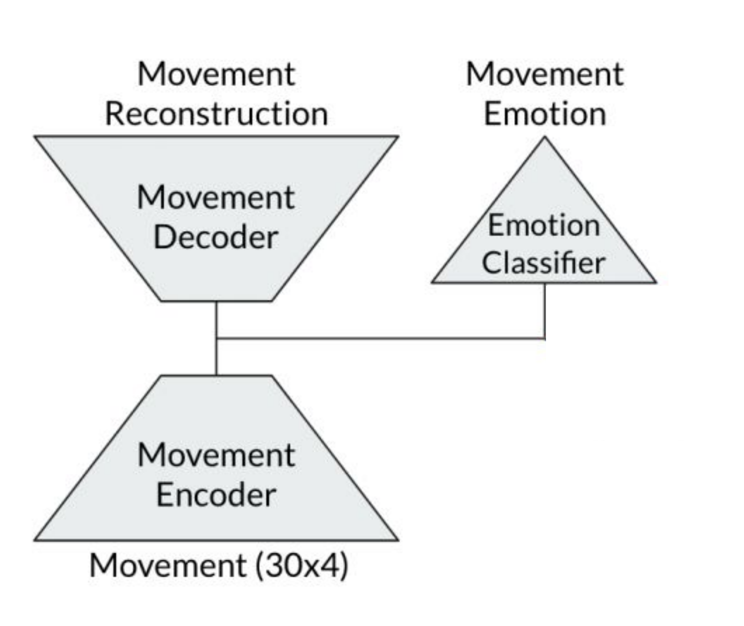
Classifier Variational Autoencoder?



Classifier Variational Autoencoder?



Classifier Variational Autoencoder?



Major Contributions

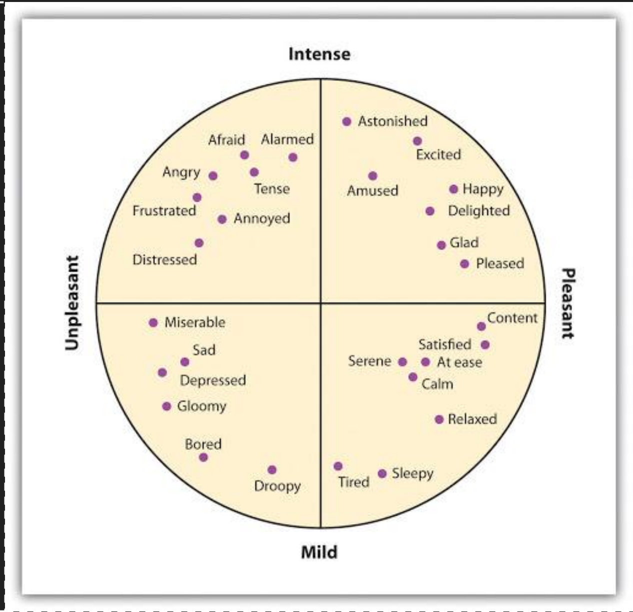
- A classifying variational autoencoder **architecture** to **reconstruct and generate** expressive **robot movements**.

- A method to **map** the latent space **representations** into the **circumplex emotional model** of valence and arousal.

- An **algorithm to modify** the valence and arousal of movements.

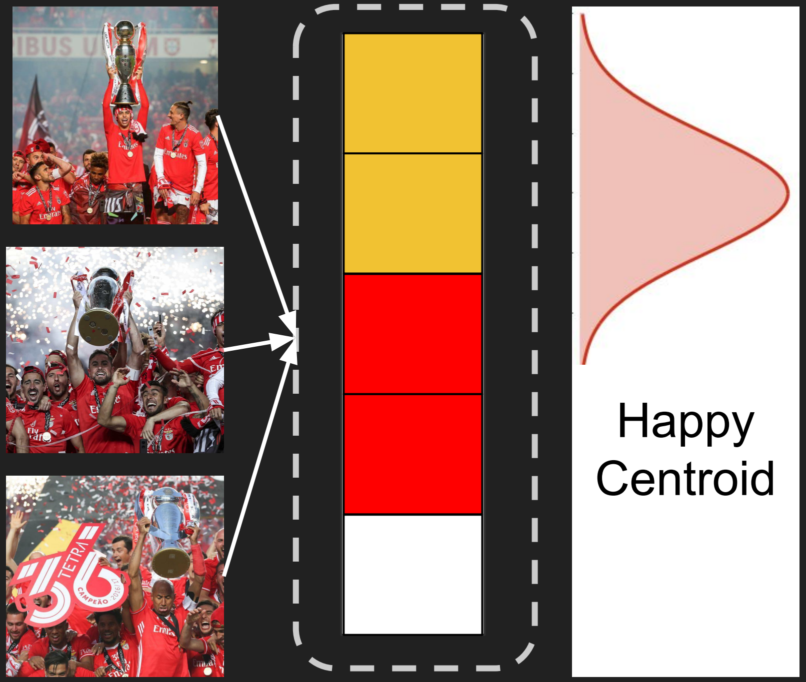
Russell's Circumplex Emotional Model (1980)

Y-axis: Arousal

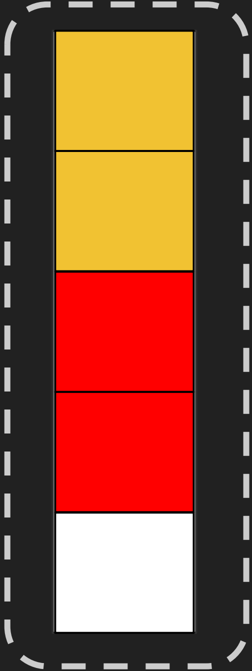


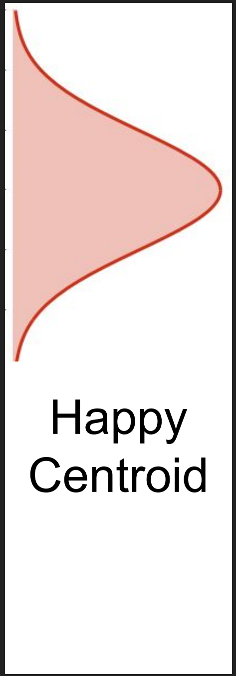
X-axis: Valence

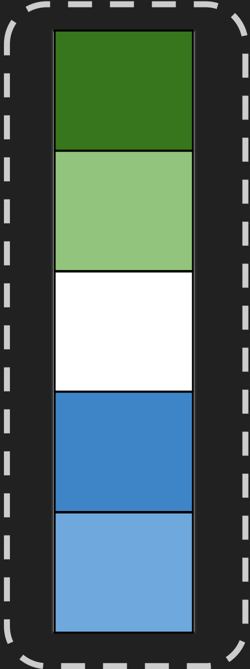
Mapping of Latent Space to Emotional Model

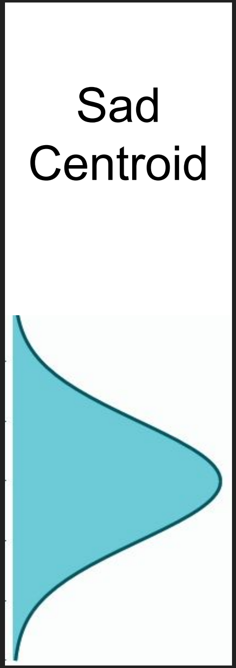


Mapping of Latent Space to Emotional Model



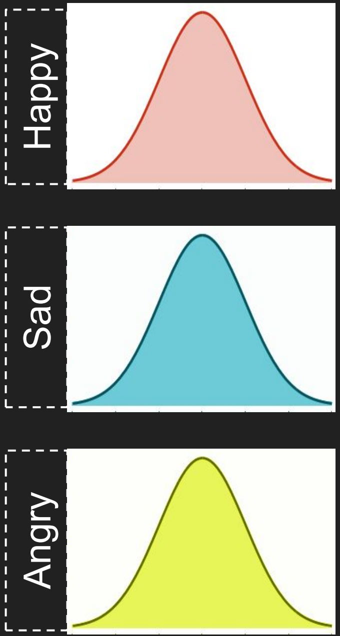






Mapping of Latent Space to Emotional Model

X = Coordinates of the centroids in Latent Space



Linear Regression

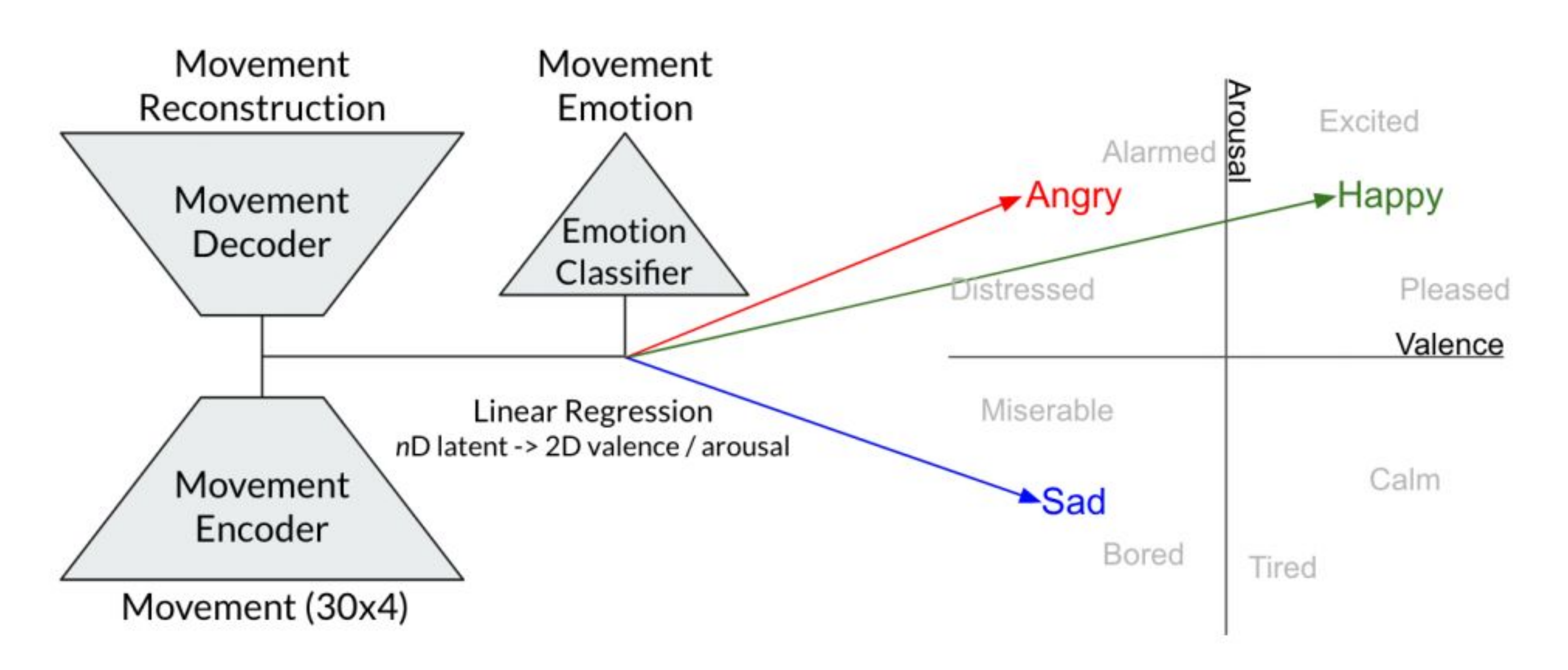
Y = Coordinates of the emotions in the circumplex model

[1, 1]

[-1, -1]

[-1, 1]

Mapping of Latent Space to Emotional Model



Major Contributions

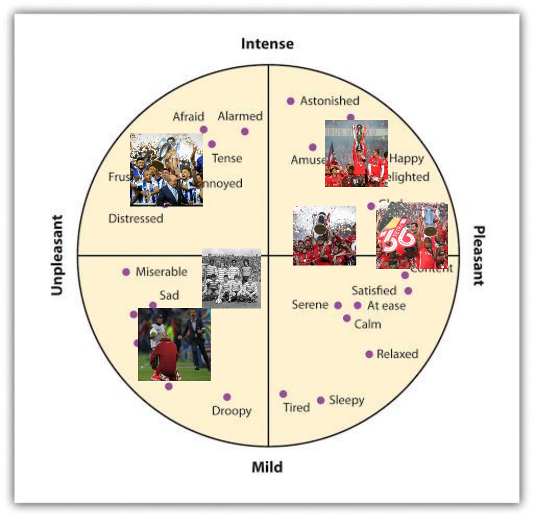
- A classifying variational autoencoder **architecture** to **reconstruct and generate** expressive **robot movements**.

- A method to **map** the latent space **representations** into the **circumplex emotional model** of valence and arousal.

- An **algorithm to modify** the valence and arousal of movements.

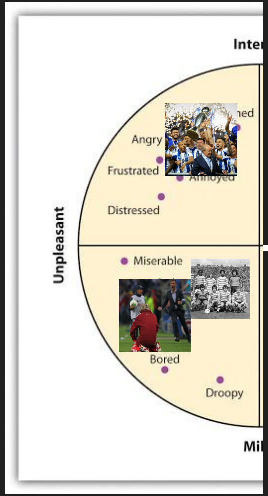
Latent Feature Modification

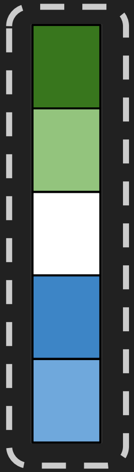
1. Rank all samples in the circumplex model, accordingly to their *valence* and *arousal* features:



Latent Feature Modification

1. For each feature *f* (Valence, Arousal), compute the average coordinates in the latent space of the lower and upper halves in the circumplex.





$$\vec{\mu}_{V,\,\mathrm{low}}$$

$$\vec{\mu}_{V,\,\mathrm{high}}$$

$$\vec{\mu}_{A,\,\mathrm{low}}$$

$$\vec{\mu}_{A,\,\mathrm{high}}$$

Latent Feature Modification

1. Compute the attribute vectors

$$\vec{a}_A = \vec{\mu}_{A,\,\mathrm{high}} - \vec{\mu}_{A,\,\mathrm{low}}$$

$$\vec{a}_V = \vec{\mu}_{V,\,\mathrm{high}} - \vec{\mu}_{V,\,\mathrm{low}}$$

Latent Feature Modification

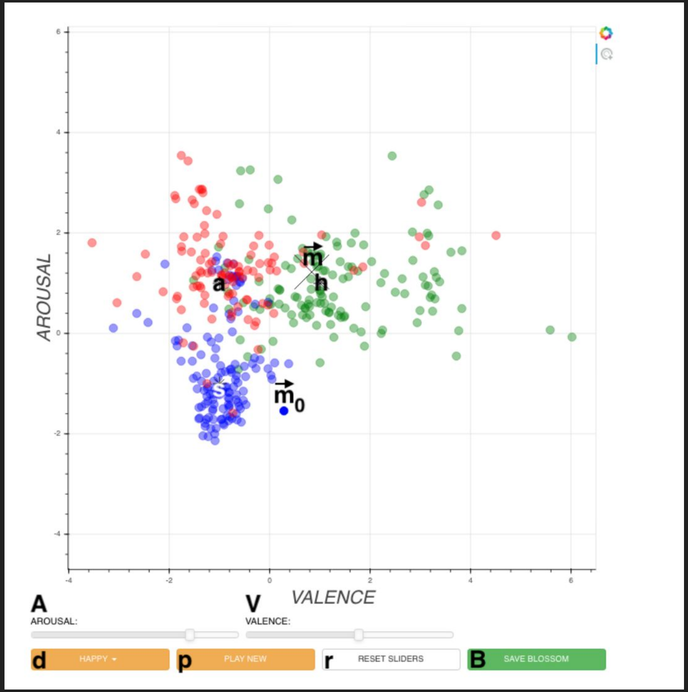
1. Modify the latent space using the attribute vectors and user-defined weights

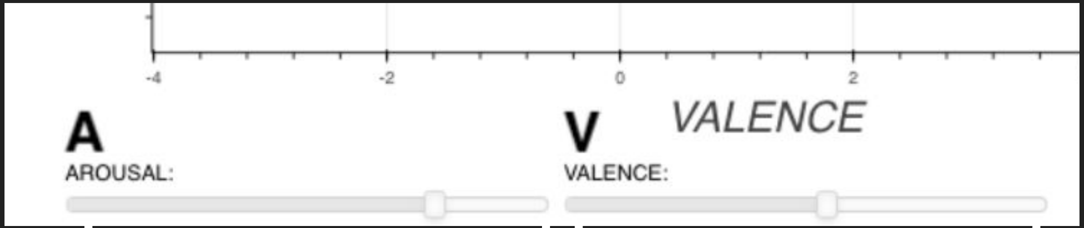
$$\vec{m} = \vec{m}_0 + c_V \vec{a}_V + c_A \vec{a}_A$$

Original movement

User Defined Weights

Latent Feature Modification





c_A c_V

Evaluation

Experimental Setup

- Robotic platform:
  - **Blossom** Robot (4 DoF)
- Dataset:
  - 3 emotion classes (Happy, Sad, Angry)
  - 25 movements per emotion class (10 Hz)
  - Chucked into 3.0 s.

The **Blossom** robot used in the evaluation of the scenario.

Experimental Setup

- Robotic platform:
  - **Blossom** Robot (4 DoF)
- Dataset:
  - 3 emotion classes (Happy, Sad, Angry)
  - 25 movements per emotion class (10 Hz)
  - Chucked into 3.0 s.
  - 5000 120-dim samples.

(3 s x 10 Hz x 4 DoFs = 120-dim)

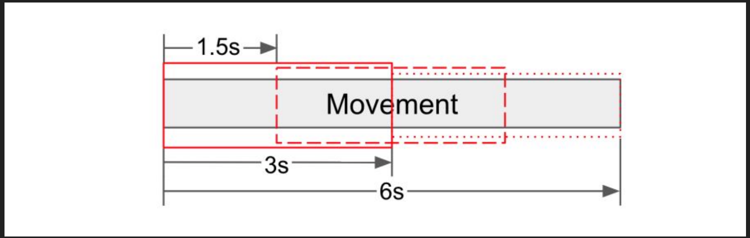


Illustration of how the movement data is "chunked" into three-second windows with 1.5-second overlaps to be used by the network.

Evaluation

- Experimental setup
- Objective metrics on the Network performance
- Subjective metrics on a online survey

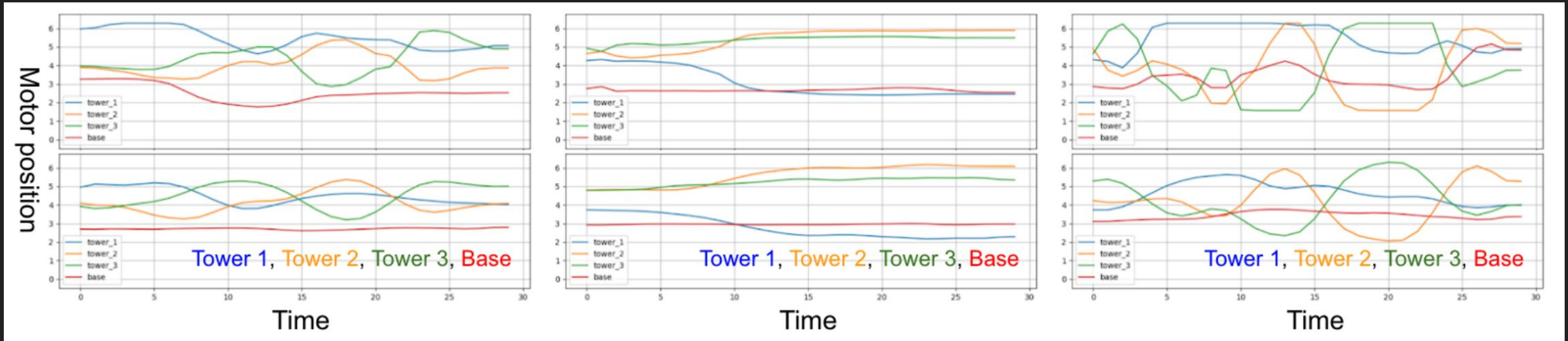
Network Performance

- **Reconstruction** capabilities:
  - "Stable training loss and validation loss close to training"...

Movement reconstruction loss over 100 epochs.

Network Performance

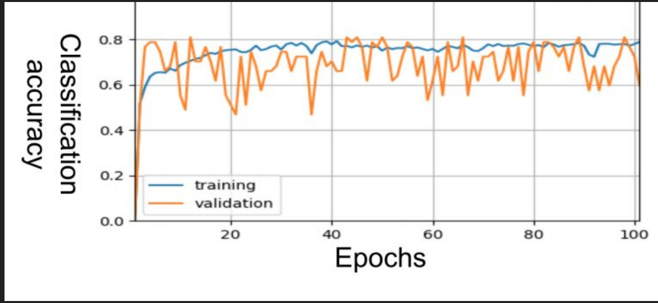
- **Reconstruction** capabilities:
  - “Reconstructed movements retain the overall trajectory characteristics”



DoF curves for original (top) and reconstructed (bottom) movements for each emotion (happy left, sad center, angry right). The blue, yellow, green, and red lines represent the front, right, left, and base motors, respectively.

Network Performance

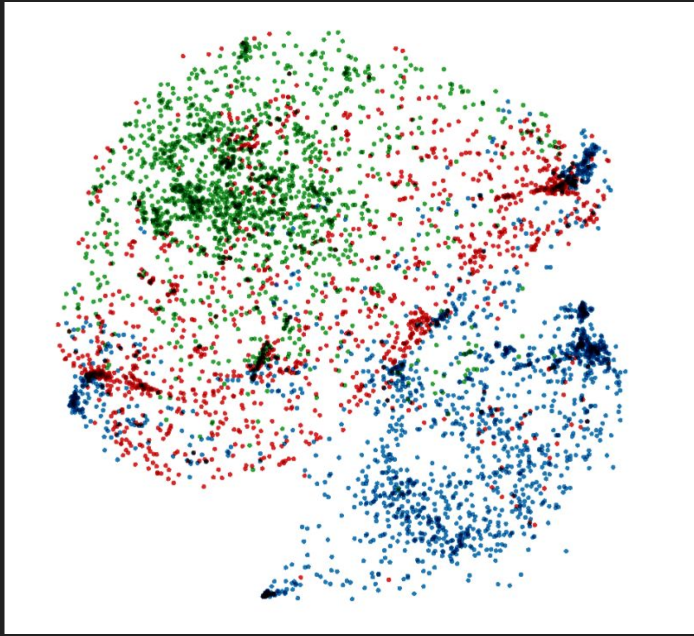
- **Classification** capabilities:
  - “Close to 80% accuracy”...



Classification accuracy over 100 epochs.

Network Performance

- **Classification** capabilities:
  - "Close to 80% accuracy"...
  - Emotion regions are separated in the latent space.



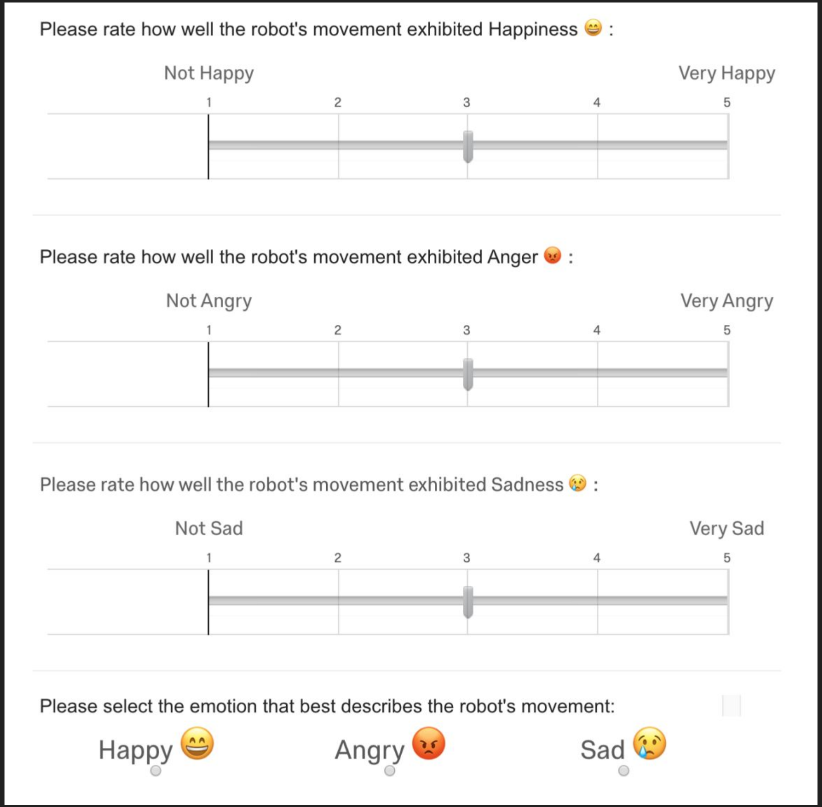
t-SNE representation of all of the movement samples in the latent space.

Evaluation

- Experimental setup
- Objective metrics on the Network performance
- Subjective metrics on a online survey

Online Survey

- Online video assessment + questionnaire
- Survey Dataset:
  - 15 original motions
  - 30 modified motions
- Every participant is shown 30 movement samples
- Each video is followed by subjective evaluation (Likert scales + multiple choice)



Online Survey

Two main hypotheses:

H1 - The modified movement's new emotion is **consistently recognized** as the target emotion.

H2 - The modified movement's new emotion expresses the target emotion **as legibly as an original movement with the same emotion**.

Online Survey - H1 results

- **TOST** (Two one-sided tests) equivalence tests (α=0,1)

Table 4: Mean emotion recognition accuracies and equivalence test *p*-values (italicized). Bolded *p*-values support H1.

		Target emotion		
		Happy	Sad	Angry
Original emotion	Happy	0.59, ——	0.63, ***0.03***	0.18, *0.13*
	Sad	0.44, *0.91*	0.66, ——	0.21, ***0.01***
	Angry	0.44, *0.91*	0.61, *0.08*	0.24, ——

- Raises questions regarding the expressiveness of the robot...

Online Survey - H2 results

- **TOST** (Two one-sided tests) equivalence tests (α=0,2)

Table 5: Mean emotion legibility scores and equivalence test *p*-values (italicized). Bolded *p*-values support H2.

		Target emotion		
		Happy	Sad	Angry
Original emotion	Happy	3.33, ——	3.42, ***0.02***	2.07, ***0.02***
	Sad	2.77, ***0.02***	3.27, ——	2.27, ***0.02***
	Angry	2.81, ***0.02***	3.54, ***0.02***	2.26, ——

- The results show that H2 is supported ($p < 0.05$) for all modifications.

Discussion Points

- Can Generative methodologies ever replace hand-animated expressive behavior? (Wish Tiago was here). Do we even want that? Would that increase the uncanny valley effect?

- Level of interpretability (legibility) between human-created and generated behavior? (Wish Silvia was here).

- How to find a balance between generated and human-provided behavior (e.g. motion, speech, etc…). Who’s to blame when something goes wrong?

Thank you all for coming!